



# Markov Property for a Function of a Markov chain: a linear algebra approach

Leonid Gurvits, James Ledoux

## ► To cite this version:

Leonid Gurvits, James Ledoux. Markov Property for a Function of a Markov chain: a linear algebra approach. Linear Algebra and its Applications, 2005, 404, pp.85-117. hal-00837503

**HAL Id: hal-00837503**

**<https://hal.science/hal-00837503>**

Submitted on 19 Aug 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Markov property for a function of a Markov chain: a linear algebra approach

Leonid Gurvits<sup>\*†</sup> and James Ledoux<sup>‡</sup>

14 april 2005

## Abstract

In this paper, we address whether a (probabilistic) function of a finite homogeneous Markov chain still enjoys a Markov-type property. We propose a complete answer to this question using a linear algebra approach. At the core of our approach is the concept of invariance of a set under a matrix. In that sense, the framework of this paper is related to the so-called “geometric approach” in control theory for linear dynamical systems. This allows us to derive a collection of new results under generic assumptions on the original Markov chain. In particular, we obtain a new criterion for a function of a Markov chain to be a homogeneous Markov chain. We provide a deterministic polynomial-time algorithm for checking this criterion. Moreover, a non-standard notion of observability for a linear system will be used. This allows one to show that the set of all stochastic matrices for which our criterion holds, is nowhere dense in the set of stochastic matrices.

**KEYWORD:**  $k$ th-order Markov chain, Rogers-Pitman matrix, Hidden Markov chain

*2000 MSC:* 15A51; 47A15; 60J10; 93B07

---

<sup>\*</sup>Los Alamos National Laboratory, New Mexico 87545, USA

<sup>†</sup>Leonid Gurvits dedicates this paper to the memory of Alexander Mihailovich Zaharin. Sasha was one of the pioneers of the Lumpability research and the most fun guy I ever knew

<sup>‡</sup>Institut National des Sciences Appliquées, 20 avenue des Buttes de Cosmes, 35043 Rennes Cedex

# 1 Introduction

Markov models are probably the most common stochastic models for dynamical systems. However, when a Markov model approach is chosen, one can expect the following issues to appear. First, computational time explosion as dimensions increase. Second, statistical properties related to a functional of the initial Markov model will be the quantities of interest, rather than the initial Markov model itself.

As a result, one either has to derive a model reduction to address the first issue, or deal with a functional of a Markov chain to address the second. Let us consider a discrete-time homogeneous Markov chain  $(X_n)$  as our initial Markov model. We assume that each random variable  $X_n$  is  $\mathcal{X}$ -valued, where  $\mathcal{X}$  is the finite set  $\{1, \dots, N\}$  and is called the *state space* of the Markov chain  $(X_n)$ . A basic way to reduce the dimension of this Markov model is to lump or collapse some states into a single “mega-state”. Thus, we obtain a partition  $\{C(1), \dots, C(M)\}$  of  $\mathcal{X}$  into  $M < N$  classes. Given such a partition, we define a map  $\varphi$  from  $\mathcal{X}$  into  $\mathcal{Y} := \{1, \dots, M\}$  by

$$\forall k \in \mathcal{X}, \forall l \in \mathcal{Y}, \quad \varphi(k) := l \iff k \in C(l).$$

The map  $\varphi$  will be referred to as a *lumping map*. Then, we are interested in the new process  $(\varphi(X_n))$  defined by

$$\varphi(X_n) = l \iff X_n \in C(l).$$

Each random variable  $\varphi(X_n)$  takes its values in the reduced set  $\mathcal{Y}$ . The process  $(\varphi(X_n))$  is called the *lumped process* with respect to the lumping map  $\varphi$ . If the original motivation is to deal with a functional of the Markov chain  $(X_n)$  then the map  $\varphi$  is obviously deduced from the functional of interest. It is well known that the lumped process  $(\varphi(X_n))$  is not Markovian in general (e.g. see [24]). Therefore, we cannot benefit from the powerful theory and algorithmic associated with the class of Markov processes. It is also known that the Markov property of  $(\varphi(X_n))$  may depend on the probability distribution of  $X_0$  which is called the *initial distribution* of the Markov chain  $(X_n)$ . If there exists an initial distribution such that  $(\varphi(X_n))$  is a homogeneous Markov chain then  $(X_n)$  is said to be *weakly lumpable*. Under specific assumptions on  $(X_n)$ , an algorithm for checking the weak lumpability property is known from [24, 35]. It is exponential in the number of states  $N$ .

We emphasize that some standard conditions for weak lumpability have been successfully apply for reducing the computational effort to deal with Markov models. These conditions are known from [24]. The most famous is the so-called *strong*

*lumpability* property:  $(\varphi(X_n))$  is a homogeneous Markov chain for every probability distribution of  $X_0$ . This has a wide range of applications in performance evaluation of computer systems (e.g. see [38, 19, 31, 9]), in control (e.g. see [14]), in chemical kinetics (e.g. see [25]), ... Another condition for weak lumpability was reported in [24]. We call it the *Rogers-Pitman condition*. An additional property makes this condition well suited for the transient assessment of large Markov chains. Indeed, the transient characteristics of the “large” Markov chain  $(X_n)$  can be obtained for computations with the “small” Markov chain  $(\varphi(X_n))$  (e.g. see [38, 10, 29, 30] and references therein). The weak lumpability property may also arise in a very general form as it is reported in [18]. The previous conditions are not satisfied but it is shown a drastic reduction of the complexity of the Markov chain resulting of a Markov chain formulation of the so-called  $k$ -SAT problem. We refer to [18] for details.

The aim of this paper is to answer to the following natural question : under which conditions, the lumped process is a  $k$ th-order Markov chain (or a non-homogeneous Markov chain). The special case  $k = 1$  corresponds to the standard homogeneous Markov property for  $(\varphi(X_n))$ . The contribution of our paper is to give a complete answer to this question as well as to some related questions, using a matrix-based approach. The significant contributions of our approach to weak lumpability are the following.

1. Up to recently, the results on weak lumpability were derived from “ergodic properties” of the Markov chain  $(X_n)$ . In this paper, we require generic assumptions on  $(X_n)$ .
2. Assume the probability distribution of  $X_0$  is fixed. We obtain a criterion for  $(\varphi(X_n))$  to be a  $k$ th order Markov chain. This condition is given in terms of some linear subspaces. This allows us to propose a deterministic polynomial-time algorithm to check this criterion.
3. We characterize the degree of freedom in the choice of the probability distribution of  $X_0$  for  $(\varphi(X_n))$  to be Markovian. This uses a concept of  $\varphi$ -*observability* which is strongly related to the standard concept of observability of a linear dynamic system.
4. If the transition matrix of the Markov chain  $(X_n)$  is  $\varphi$ -observable, then the Rogers-Pitman condition is essentially a criterion for weak lumpability property to hold.

5. It has been reported in the literature that, in practice, one rarely encounters a Markovian function of a Markov model. We provide a quantitative assessment on this fact. We show that the set of stochastic matrices for which our criterion for weak lumpability is satisfied, is nowhere dense in the set of the stochastic matrices.

A last contribution concerns the so-called probabilistic functions of Markov chains as defined by Petrie [7]. This class of Markov models is referred to as the class of *hidden Markov chains* in a modern terminology. The hidden Markov chains has a wide range of applications in time series modeling (see [15] for a recent review). Since this class of Markov models may be placed in the context discussed above, all previously mentioned results may be applied to derive specific results on the lumpability of hidden Markov chains. To the best of our knowledge, the lumpability of hidden Markov chains has received attention only in two recent papers, in view of reducing the complexity of filtering [37, 42]. These results are corollaries of the results presented here. We only consider in this paper the problem discussed in [37]: under which conditions does the observed process associated with a hidden Markov chain have the Markov property?

The framework of this paper is essentially based on the concept of sets invariant under a matrix. The readers which are familiar with the so-called “geometric approach” in linear systems theory, will find similarity between the linear subspaces introduced in this approach and those used in this paper (see [6, 43] and the references therein). The connection will be made explicit throughout the paper. The interplay between standard “lumping” questions and the “geometric approach” was one the topics in [20]. The main “geometric” framework of this paper, including a criterion for weak lumpability, was first discovered by the first author in early 1980-85 [20]. The second author had discovered independently an almost similar approach about ten years later [27, 28].

Now, we briefly discuss some topics which may appear to be related to our work.

The question of *aggregation of variables* of linear dynamic systems is connected to the questions examined in the paper. Some attention has been given to the following rather easy problem (e.g. see [39]): given a deterministic input-output linear dynamic system

$$x_{n+1} = Ax_n, \quad y_n = Cx_n$$

under which conditions does the sequence  $(y_n)$  have a linear dynamics? This question is relevant to the Markov framework when  $x_n$  is the probability distribution of the random variable  $X_n$ ,  $A$  is the transition matrix of the Markov chain  $(X_n)$  and the matrix  $C$  specifies the lumping map  $\varphi$ . In this case,  $y_n$  is the probability

distribution of the random variable  $\varphi(X_n)$ . Hence, a linear dynamic for the sequence  $(y_n)$  is a property of one-dimensional distributions of the process  $(\varphi(X_n))$ . In our paper, the Markov property for  $(\varphi(X_n))$  is a property of the collection of all the finite-dimensional distributions of this process. At times, this difference has been overlooked because the two problems have the same answer under the additional requirement that the sequence  $(y_n)$  has a linear dynamic for every stochastic vector  $x_0$  (i.e. for every probability distribution of  $X_0$ ). A discussion about aggregation of variables in the Markov framework is reported in the recent paper [30].

The problem of identification of models has a strong connection with our work. Indeed, the problem is to determine whether functions of two Markov chains give rise to the same stochastic process. This will be apparent in Section 3, when we use the “non- $\varphi$ -observable subspace” introduced in [3]. In contrast, we would like to mention that the question of *stochastic realization* is only weakly connected to our problem (e.g. see [32, 1] and the references therein). Indeed, the closest formulation of the stochastic realization problem to our setting is: given a stochastic process  $(Y_n)$ , are there a function  $\varphi$  and a Markov chain  $(X_n)$  such that the stochastic processes  $(Y_n)$  and  $(\varphi(X_n))$  have the same finite-dimensional distributions. Our problem is not to find such a Markovian representation or realization of  $(Y_n)$ . In our setting, the process  $(X_n)$  and  $\varphi$  are given. However, we mention that a minimal realization of the process  $(\varphi(X_n))$  may be obtained from the concepts of  $\varphi$ -observability and strong lumpability discussed here [22].

The paper is organized as follows. We begin by introducing the basic notations and conventions used throughout this paper. In Section 2, we propose a complete study of the Markov property for a deterministic function of  $(X_n)$ . In Subsection 2.1, we prove a criterion for  $(\varphi(X_n))$  to be a  $k$ th-order Markov chain. Under a non-singularity assumption for some blocks of the transition matrix of  $(X_n)$ , the  $k$ th-order Markov property ( $k \geq 2$ ) and the usual weak lumpability property ( $k = 1$ ) are shown to be equivalent. In Subsection 2.2, we specialize the previous results to the order one, which corresponds to the usual weak lumpability. This gives our main criterion for weak lumpability. We also give a new sufficient condition for weak lumpability to hold. Next, we outline a deterministic polynomial-time algorithm to check weak lumpability. At this step, we briefly discuss the computation of the set of all initial distributions for which  $(\varphi(X_n))$  is an Markov chain with a transition matrix that does not depends on the initial distribution. In Subsection 2.2.4, we relate the weak lumpability property of  $(X_n)$  to that of its “reversed” or “dual” version. In particular, we prove that weak and strong lumpability properties coincide when  $(X_n)$  has an irreducible and normal transition matrix. In the last part of Section 2, we deal with the non-homogeneous Markov property of  $(\varphi(X_n))$ . We

obtain a "nice" answer only in the periodic case, that is, when the sequence of transition matrices of  $(\varphi(X_n))$  is periodic. In Section 3, we present the concept of  $\varphi$ -observability. In Subsection 3.1, it is shown that under  $\varphi$ -observability, the Rogers-Pitman condition becomes essentially a criterion for weak lumpability property to hold. In Subsection 3.2, the set of all weakly lumpable matrices is shown to be nowhere dense in the set of stochastic matrices. We turn to the Markov property for the observed process of a hidden Markov chain in Section 4. Basic criteria are stated in terms of the standard parameters of such processes. We conclude in Section 5.

## Preliminaries

- A vector is a column vector by convention.  $()^T$  denotes the transpose operator. The  $i$ th component of a vector  $u$  is denoted by  $u(i)$ . Any inequality between vectors is understood as being component-wise.
- $\mathbf{1}$  (resp.  $\mathbf{0}$ ) stands for a finite-dimensional vector with each component equals to 1 (resp. 0). Its dimension is defined by the context.
- $\mathcal{X}$ ,  $\mathcal{Y}$  denote the finite sets  $\{1, \dots, N\}$  and  $\{1, \dots, M\}$  respectively, with  $M < N$ .
- $P$  denotes a  $N \times N$  stochastic matrix, i.e.  $P$  is a non-negative matrix such that  $\mathbf{1}^T P = \mathbf{1}^T$ .  $\hat{I}$  is the  $M \times M$  identity matrix.
- The random elements are assumed to be defined on the same probability space with probability  $(\Omega, \mathcal{F}, \mathbb{P})$ , where  $\mathbb{P}$  is a probability on the  $\sigma$ -algebra  $\mathcal{F}$  of events.
- The probability distribution of a  $\mathcal{X}$ -valued (resp  $\mathcal{Y}$ -valued) random variable  $W$  will be identified with a  $N$ -dimensional (resp.  $M$ -dimensional) stochastic vector  $\alpha$  defined by  $\alpha(x) := \mathbb{P}\{W = x\}, x \in \mathcal{X}$  (resp.  $x \in \mathcal{Y}$ ).
- A set  $\mathcal{C}$  is said to be *invariant under a matrix*  $A$  if

$$A \mathcal{C} \subset \mathcal{C},$$

where  $A \mathcal{C} := \{A c : c \in \mathcal{C}\}$ . We also say that  $\mathcal{C}$  is  $A$ -invariant.

- Let  $\varphi$  be a non-decreasing map from  $\mathcal{X}$  into  $\mathcal{Y}$  such that  $\varphi(\mathcal{X}) = \mathcal{Y}$ .

- We define a  $M \times N$  matrix  $V_\varphi$  by

$$V_\varphi(y, x) := 1 \text{ if } x \in \varphi^{-1}(y) \quad \text{and} \quad 0 \text{ otherwise.}$$

- For any  $y \in \mathcal{Y}$ , let  $\Pi_y$  be the  $N \times N$  matrix defined by

$$\Pi_y(x, x) := 1 \text{ if } x \in \varphi^{-1}(y) \quad \text{and} \quad \Pi_y(x_1, x_2) := 0 \text{ otherwise.}$$

This matrix will be called the *y-lumping projector*.

For any  $v \in \mathbb{R}^N$ ,  $V_\varphi v$  is the  $M$ -dimensional vector

$$V_\varphi v = (\mathbf{1}^T \Pi_y v)_{y \in \mathcal{Y}}. \quad (1.1)$$

We mention that the kernel  $\text{Ker}(V_\varphi)$  is  $\Pi_y$ -invariant for any  $y \in \mathcal{Y}$ , that is

$$\text{Ker}(V_\varphi) = \oplus_{y \in \mathcal{Y}} \Pi_y \text{Ker}(V_\varphi).$$

- Let  $\mathcal{Y}^*$  be the set of all finite sequences of elements in  $\mathcal{Y}$ . For any  $s \in \mathcal{Y}^*$ ,  $P^{(s)}$  is the  $N \times N$  matrix defined by

$$P^{(s)} := \begin{cases} \Pi_{y_0} & \text{if } s = y_0 \\ \Pi_{y_n} P P^{(y_{n-1} \dots y_0)} & \text{if } s = y_n \dots y_1 y_0. \end{cases} \quad (1.2)$$

For  $s = y_n \dots y_1 y_0$ , we set  $\text{lg}(s) := n$ .

- For any non-negative  $N$ -dimensional vector  $v$  such that  $\mathbf{1}^T \Pi_y v > 0$ ,  $v^{(y)}$  is the  $N$ -dimensional vector

$$v^{(y)} := \frac{\Pi_y v}{\mathbf{1}^T \Pi_y v}.$$

- For any non-negative vector  $v$  such that  $V_\varphi v > \mathbf{0}$ , the matrix  $U_v$ , defined by

$$x \in \mathcal{X}, y \in \mathcal{Y}, \quad U_v(x, y) := v^{(y)}(x),$$

is such that  $V_\varphi U_v = \hat{I}$ . An equality between matrices of the type  $U_v$  with the mention “(if well-defined)” means that the equality holds when the matrices and the vectors in the under-script of these matrices are well-defined according to our definitions.



- The following definitions are borrowed from [8]. For any subset  $\mathcal{C}$  of  $\mathbb{R}^n$ ,  $\text{Span}(\mathcal{C})$  (resp.  $\text{Cone}(\mathcal{C})$ ) denotes the set of all finite (resp. non-negative) linear combinations of the elements of  $\mathcal{C}$ . If  $\mathcal{C}$  is a set of non-negative vectors, then  $\text{Span}(\mathcal{C}) = \text{Cone}(\mathcal{C}) - \text{Cone}(\mathcal{C})$ . If  $\text{Cone}(\mathcal{C}) = \mathcal{C}$  then  $\mathcal{C}$  is called a *cone*. If  $\mathcal{C}$  is a finite set, then  $\text{Cone}(\mathcal{C})$  is called a *polyhedral cone*. Any polyhedral cone  $\mathcal{C}$  of  $\mathbb{R}^n$  has the form  $\mathcal{C} = \{v \in \mathbb{R}^n : Hv \geq \mathbf{0}\}$  where  $H \in \mathbb{R}^{m \times n}$ . This is a closed convex subset of  $\mathbb{R}^n$ .

The cone  $\mathcal{C}$  is said to be *decomposable* if  $\mathcal{C} = \mathcal{C}_1 + \mathcal{C}_2$ , where  $\mathcal{C}_1$  and  $\mathcal{C}_2$  are two sub-cones of  $\mathcal{C}$  such that  $\text{Span}(\mathcal{C}_1) \cap \text{Span}(\mathcal{C}_2) = \{\mathbf{0}\}$ . We write  $\mathcal{C} = \mathcal{C}_1 \oplus \mathcal{C}_2$ .

Our basic instance of a decomposable cone is the cone denoted by  $\mathcal{CC}(\alpha, \Pi, P)$  for a fixed  $N$ -dimensional stochastic vector  $\alpha$ . It is defined as the smallest sub-cone of  $\mathbb{R}_+^N$  that contains the vector  $\alpha$  and that is invariant under the matrix  $P$  and the lumping projectors  $\Pi_y, y \in \mathcal{Y}$ . It is easily seen that

$$\mathcal{CC}(\alpha, \Pi, P) := \text{Cone}(P^{(s)}\alpha, s \in \mathcal{Y}^*). \quad (1.3)$$

The basic properties of this cone are from its definition

$$y \in \mathcal{Y}, \quad \Pi_y \mathcal{CC}(\alpha, \Pi, P) \subset \mathcal{CC}(\alpha, \Pi, P) \iff \mathcal{CC}(\alpha, \Pi, P) = \oplus_{y \in \mathcal{Y}} \Pi_y \mathcal{CC}(\alpha, \Pi, P) \\ P \mathcal{CC}(\alpha, \Pi, P) \subset \mathcal{CC}(\alpha, \Pi, P).$$

- For a  $N$ -dimensional stochastic vector  $\alpha$ , the central linear subspace used in this paper, is the linear hull  $\mathcal{CS}(\alpha, \Pi, P)$  of cone  $\mathcal{CC}(\alpha, \Pi, P)$  defined above. That is

$$\mathcal{CS}(\alpha, \Pi, P) := \mathcal{CC}(\alpha, \Pi, P) - \mathcal{CC}(\alpha, \Pi, P) = \text{Span}(P^{(s)}\alpha, s \in \mathcal{Y}^*). \quad (1.4)$$

The subspace  $\mathcal{CS}(\alpha, \Pi, P)$  is the minimal subspace that contains  $\alpha$  and that is invariant under  $P$  and the lumping projectors  $\Pi_y, y \in \mathcal{Y}$ . In particular, it satisfies

$$y \in \mathcal{Y}, \quad \Pi_y \mathcal{CS}(\alpha, \Pi, P) \subset \mathcal{CS}(\alpha, \Pi, P) \iff \mathcal{CS}(\alpha, \Pi, P) = \oplus_{y \in \mathcal{Y}} \Pi_y \mathcal{CS}(\alpha, \Pi, P) \\ P \mathcal{CS}(\alpha, \Pi, P) \subset \mathcal{CS}(\alpha, \Pi, P).$$

## 2 Markov property for a function of an HMC

The Markov chains are the basic stochastic processes considered in this paper. We recall the definition of a  $k$ th-order Markov chain. For  $k = 1$ , the process is simply said to be a homogeneous Markov chain.

**Definition 2.1** Let  $\mathcal{E}$  be a finite set. A sequence of  $\mathcal{E}$ -valued random variables  $(Z_n)$  is said to be a homogeneous  $k$ th-order Markov chain ( $k$ th-HMC) with the transition matrix  $Q$  iff for any  $n \geq 0$ ,  $(z, z_{n+k-1}, \dots, z_0) \in \mathcal{E}^{n+k+1}$ ,

$$\mathbb{P}\{Z_{n+k} = z \mid Z_{n+k-1} = z_{n+k-1}, \dots, Z_n = z_n, \dots, Z_0 = z_0\} = Q(z, z_{n+k-1}, \dots, z_n) \\ \text{if } \mathbb{P}\{Z_{n+k-1} = z_{n+k-1}, \dots, Z_0 = z_0\} > 0.$$

The set  $\mathcal{E}$  is called the state space of  $(Z_n)$ . Note that  $\sum_{z \in \mathcal{E}} Q(z, z_{n+k-1}, \dots, z_n) = 1$ . The probability distribution of  $Z_0$  is called the initial distribution of the process  $(Z_n)$ .

Let  $(X_n)_n$  be an HMC with state space  $\mathcal{X}$ . In Section 2, the  $N \times N$  matrix  $P$  and the  $N$ -dimensional stochastic vector  $\alpha$  will stand for the transition matrix and the initial distribution of the Markov chain  $(X_n)$  respectively. Let  $\varphi$  be a non-decreasing map of  $\mathcal{X}$  onto  $\mathcal{Y}$  such that

$$\varphi(\mathcal{X}) = \mathcal{Y}, \quad M < N.$$

Such a map defines a partition of  $\mathcal{X}$  into the  $M$  classes  $(\varphi^{-1}(y), y \in \mathcal{Y})$  and

$$\varphi(X_n) = y \iff X_n \in \varphi^{-1}(y).$$

The function  $\varphi$  will be called a *lumping map*. The process  $(\varphi(X_n))$  will be called the *lumped process* associated with  $(X_n)_n$ . Considering the string  $s := y_n \dots y_1 y_0 \in \mathcal{Y}^*$  means that we are interested in the successive visited mega-states  $y_0, \dots, y_n$  by the lumped model. The integer  $\text{lg}(s)$  corresponds to the number of transitions in the path  $y_0, \dots, y_n$ . Note that with (1.2)

$$\mathbb{P}\{\varphi(X_n) = y_n, \dots, \varphi(X_0) = y_0\} = \mathbf{1}^T \Pi_{y_n} P \Pi_{y_{n-1}} \dots \Pi_{y_1} P \Pi_{y_0} \alpha = \mathbf{1}^T P^{(s)} \alpha. \quad (2.1)$$

## 2.1 Homogeneous $k$ th-order Markov property

The following theorem provides a first criterion for  $(\varphi(X_n))$  to be a  $k$ th-HMC. This criterion is in terms of the cone  $\mathcal{CC}(\alpha, \Pi, P)$  defined in (1.3).

**Theorem 2.1** The process  $(\varphi(X_n))$  is a  $k$ th-HMC with transition matrix  $\hat{P}$  iff

$$\mathcal{CC}(\alpha, \Pi, P) \subset \mathcal{C}_k(\hat{P}),$$

where

$$\mathcal{C}_k(\hat{P}) := \{\beta \geq 0 : \forall (y_{k-1}, \dots, y_0) \in \mathcal{Y}, [V_\varphi P - \hat{P}_{y_{k-1}, \dots, y_0} \mathbf{1}^T] P^{(y_{k-1} \dots y_0)} \beta = 0\} \quad (2.2)$$

and  $\widehat{P}_{y_{k-1}, \dots, y_0}$  denotes the vector  $(\widehat{P}(y, y_{k-1}, \dots, y_0))_{y \in \mathcal{Y}}$ .

If the above inclusion holds, then  $(\varphi(X_n))$  is still a  $k$ th-HMC with any stochastic vector in  $\mathcal{CC}(\alpha, \Pi, P)$  as initial distribution of  $(X_n)$ .

*Proof.* Let  $\widetilde{\mathcal{Y}}^*$  be the set  $\mathcal{Y}^*$  complemented by the empty sequence. From Definition 2.1 and (2.1),  $(\varphi(X_n))$  is a  $k$ th-HMC with transition matrix  $\widehat{P}$  iff for any  $(y, y_{k-1}, \dots, y_0) \in \mathcal{Y}^{k+1}$  and  $s \in \widetilde{\mathcal{Y}}^*$

$$\mathbf{1}^T P^{(yy_{k-1} \dots y_0 s)} \alpha = \widehat{P}(y; y_{k-1}, \dots, y_0) \mathbf{1}^T P^{(y_{k-1} \dots y_0 s)} \alpha.$$

Since  $P^{(yy_{k-1} \dots y_0 s)} = \Pi_y P P^{(y_{k-1} \dots y_0 s)}$  (see (1.2)), the previous statement has the following equivalent form from (1.1)

$$\begin{aligned} \forall (y_{k-1}, \dots, y_0) \in \mathcal{Y}^k, \forall s \in \widetilde{\mathcal{Y}}^* \quad & V_\varphi P P^{(y_{k-1} \dots y_0 s)} \alpha = \widehat{P}_{y_{k-1}, \dots, y_0} \mathbf{1}^T P^{(y_{k-1} \dots y_0 s)} \alpha \\ \iff \forall (y_{k-1}, \dots, y_0) \in \mathcal{Y}^k, \forall s \in \widetilde{\mathcal{Y}}^* \quad & [V_\varphi P - \widehat{P}_{y_{k-1}, \dots, y_0} \mathbf{1}^T] P^{(y_{k-1} \dots y_0 s)} \alpha = \mathbf{0} \\ \iff \forall (y_{k-1}, \dots, y_0) \in \mathcal{Y}^k, \forall s \in \widetilde{\mathcal{Y}}^* \quad & [V_\varphi P - \widehat{P}_{y_{k-1}, \dots, y_0} \mathbf{1}^T] P^{(y_{k-1} \dots y_0)} P^{(y_0 s)} \alpha = \mathbf{0} \\ \iff \forall y_0 \in \mathcal{Y}, \forall s \in \widetilde{\mathcal{Y}}^* \quad & P^{(y_0 s)} \alpha \in \mathcal{C}_k(\widehat{P}) \quad \text{from (2.2)} \\ \iff \forall s \in \mathcal{Y}^* \quad & P^{(s)} \alpha \in \mathcal{C}_k(\widehat{P}). \end{aligned}$$

That  $P^{(s)} \alpha \in \mathcal{CC}(\alpha, \Pi, P)$  for any  $s \in \mathcal{Y}^*$ , is clear from (1.3). Then, we have  $\mathcal{CC}(P^{(s)} \alpha, \Pi, P) \subset \mathcal{CC}(\alpha, \Pi, P)$  and the last statement of the theorem follows.  $\blacksquare$

We derive now a criterion for  $(\varphi(X_n))$  to be an  $k$ th-HMC in terms of the linear subspace  $\mathcal{CS}(\alpha, \Pi, P)$  defined in (1.4). The knowledge of the matrix  $\widehat{P}$  is not required.

**Theorem 2.2** *The process  $(\varphi(X_n))$  is a  $k$ th-HMC iff*

$$\forall (y_{k-1}, \dots, y_0) \in \mathcal{Y}^k, \quad P(\text{Ker}(V_\varphi) \cap P^{(y_{k-1} \dots y_0)} \mathcal{CS}(\alpha, \Pi, P)) \subset \text{Ker}(V_\varphi). \quad (2.3)$$

If Assertion (2.3) holds then  $(\varphi(X_n))$  is still a  $k$ th-HMC with any stochastic vector in  $\mathcal{CS}(\alpha, \Pi, P)$  as initial distribution of  $(X_n)$ .

The connections with the “geometric approach” in linear system theory is clear from (1.4) and (2.3). That the subspace  $\mathcal{CS}(\alpha, \Pi, P)$  is the minimal subspace that is invariant under  $P$  and the lumping projectors  $\Pi_y, y \in \mathcal{Y}$  and that contains the subspace  $\text{Span}(\Pi_y \alpha, y \in \mathcal{Y})$ , is easily seen from its definition. A specific algorithm for computing the subspace  $\mathcal{CS}(\alpha, \Pi, P)$  may be easily designed from [6]. We turn to this question in Subsection 2.2.2. Note that the subspace  $\text{Ker}(V_\varphi)$  is a

$(P, P^{(y_{k-1} \dots y_0)} \mathcal{CS}(\alpha, \Pi, P))$ -conditioned invariant according to the terminology used by Basile and Marro [6]. This last fact is not useful for our purpose.

*Proof* Let  $(\varphi(X_n))$  be a  $k$ th-HMC with some transition matrix  $\hat{P}$ . We know from Theorem 2.1 that  $\mathcal{CC}(\alpha, \Pi, P) \subset \mathcal{C}_k(\hat{P})$ . Any vector  $v \in P^{(y_{k-1} \dots y_0)} \mathcal{CS}(\alpha, \Pi, P)$  has the form  $P^{(y_{k-1} \dots y_0)}(\beta - \gamma)$  with  $\beta, \gamma \in \mathcal{C}_k(\hat{P})$ . If  $v \in \text{Ker}(V_\varphi)$  then

$$\begin{aligned} V_\varphi P v &= V_\varphi P P^{(y_{k-1} \dots y_0)}(\beta - \gamma) \\ &= \hat{P}_{y_{k-1}, \dots, y_0} \mathbf{1}^T P^{(y_{k-1} \dots y_0)}(\beta - \gamma) \quad \text{since } \beta, \gamma \in \mathcal{C}_k(\hat{P}) \\ &= \hat{P}_{y_{k-1}, \dots, y_0} \mathbf{1}^T v = \mathbf{0} \quad \text{since } \mathbf{1}^T v = 0. \end{aligned}$$

Thus, the vector  $Pv$  is in  $\text{Ker}(V_\varphi)$ .

Assume we have the inclusion in (2.3). In a first step, we define the matrix  $\hat{P}$  as follows. For any  $(y_{k-1}, \dots, y_0) \in \mathcal{Y}^k$  such that  $P^{(y_{k-1} \dots y_0)} \mathcal{CC}(\alpha, \Pi, P) \neq \{\mathbf{0}\}$ , select a non-trivial vector  $\beta$  in this set and put

$$\hat{P}_{y_{k-1} \dots y_0} := V_\varphi P \beta^{(y_{k-1})}. \quad (2.4)$$

When  $P^{(y_{k-1} \dots y_0)} \mathcal{CC}(\alpha, \Pi, P) = \{\mathbf{0}\}$ , it is easily seen that, with  $\mathbb{P}$ -probability 1, the path  $y_{k-1} \dots y_0$  is not observable for  $(\varphi(X_n))$ . Then, the stochastic vector  $\hat{P}_{y_{k-1} \dots y_0}$  may be arbitrary chosen.

In a second step, we have to prove that  $\mathcal{CC}(\alpha, \Pi, P) \subset \mathcal{C}_k(\hat{P})$ . If  $\mathcal{CC}(\alpha, \Pi, P) = \{\mathbf{0}\}$  then this is trivially true. When  $\mathcal{CC}(\alpha, \Pi, P) \neq \{\mathbf{0}\}$ , let  $\gamma$  be a non-trivial vector of  $\mathcal{CC}(\alpha, \Pi, P)$ . We must find that

$$[V_\varphi P - \hat{P}_{y_{k-1}, \dots, y_0} \mathbf{1}^T] P^{(y_{k-1} \dots y_0)} \gamma = \mathbf{0}. \quad (2.5)$$

If  $P^{(y_{k-1} \dots y_0)} \gamma = \mathbf{0}$ , this is obvious. If  $P^{(y_{k-1} \dots y_0)} \gamma \neq \mathbf{0}$ , then Equation (2.5) has the following equivalent form from (2.4)

$$V_\varphi P(P^{(y_{k-1} \dots y_0)} \gamma)^{(y_{k-1})} = \hat{P}_{y_{k-1} \dots y_0} = V_\varphi P \beta^{(y_{k-1})}.$$

Since  $(P^{(y_{k-1} \dots y_0)} \gamma)^{(y_{k-1})} - \beta^{(y_{k-1})} \in \text{Ker}(V_\varphi)$ , we have  $(P^{(y_{k-1} \dots y_0)} \gamma)^{(y_{k-1})} - \beta^{(y_{k-1})} \in \text{Ker}(V_\varphi) \cap P^{(y_{k-1} \dots y_0)} \mathcal{CS}(\alpha, \Pi, P)$ . We find from the inclusion (2.3) that

$$\mathbf{0} = V_\varphi P[(P^{(y_{k-1} \dots y_0)} \gamma)^{(y_{k-1})} - \beta^{(y_{k-1})}] \iff V_\varphi P(P^{(y_{k-1} \dots y_0)} \gamma)^{(y_{k-1})} = V_\varphi P \beta^{(y_{k-1})}.$$

■

**Corollary 2.1** Suppose the matrix  $(P(x_1, x_2))_{x_1, x_2 \in \varphi^{-1}(y)}$  is non-singular for every  $y \in \mathcal{Y}$ . If  $(\varphi(X_n))$  is an  $k$ th-HMC for some  $k \geq 2$  then  $(\varphi(X_n))$  is an HMC.

*Proof.* f  $(P(x_1, x_2))_{x_1, x_2 \in \varphi^{-1}(y)}$  is a non-singular matrix then

$$\forall \mathcal{Z} \subset \mathbb{R}^N, \forall y \in \mathcal{Y}, \quad P^{(y \dots y)} \mathcal{Z} = P^{(yy)} \mathcal{Z} = \Pi_y \mathcal{Z}.$$

If  $(\varphi(X_n))$  is a  $k$ th-HMC, then we deduce from (2.3) with  $y_{k-1} \dots y_0 := y \dots y$  and the equality above, that

$$\forall y \in \mathcal{Y}, \quad P(\text{Ker}(V_\varphi) \cap \Pi_y \mathcal{CS}(\alpha, \Pi, P)) \subset \text{Ker}(V_\varphi).$$

Thus,  $(\varphi(X_n))$  is a 1th-HMC from Theorem 2.2.  $\blacksquare$

We mention for completeness, the following spectral properties resulting from the  $k$ th-order lumpability property. They hold because the cone  $\mathcal{CC}(\alpha, \Pi, P)$  is invariant under the matrices  $P$  and  $\Pi_y P \Pi_y$ ,  $y \in \mathcal{Y}$  (see [27, Lemma 3.3]). When a probabilistic approach is favored, these properties combined with the additional assumption of the existence of an “ergodic” type theorem, allow the derivation of results on (1-order) lumpability through limit arguments (e.g. see [24, 35]).

### Corollary 2.2

1. If  $(\varphi(X_n))$  is a  $k$ th-HMC, then it is still a  $k$ th-HMC with some stochastic eigenvector of  $P$  as initial distribution of  $(X_n)$ .
2. If  $\alpha$  is such that  $\Pi_y \alpha \neq \mathbf{0}$  and  $(\varphi(X_n))$  is a  $k$ th-HMC, then  $(\varphi(X_n))$  is still a  $k$ th-HMC with some stochastic eigenvector of  $\Pi_y P \Pi_y$  as initial distribution of  $(X_n)$ .

**Comment 1** If  $P$  is irreducible then the Perron-Frobenius theorem asserts that there exists an unique stochastic eigenvector  $\pi$  corresponding to the eigenvalue 1. This vector is the stationary distribution of the HMC  $(X_n)$ . Let  $\text{distr}(X_0)$  denotes the probability distribution of  $X_0$ . Then, we can write

$$\begin{aligned} & (\varphi(X_n)) \text{ is a } k\text{th-HMC (with transition matrix } \widehat{P}) \text{ for } \text{distr}(X_0) := \alpha \\ \implies & (\varphi(X_n)) \text{ is a } k\text{th-HMC (with transition matrix } \widehat{P}) \text{ for } \text{distr}(X_0) := \pi \\ \implies & (\varphi(X_n)) \text{ is a } k\text{th-HMC (with transition matrix } \widehat{P}) \text{ for } \text{distr}(X_0) \in \mathcal{C}_\pi \\ & \text{with } \mathcal{C}_\pi := \bigoplus_{y \in \mathcal{Y}} \text{Cone}(\Pi_y \pi) \end{aligned}$$

since  $\mathcal{CC}(u, \Pi, P) = \mathcal{CC}(\pi, \Pi, P)$  for any  $u \in \mathcal{C}_\pi$ .  $\triangle$

Let  $\mathcal{Z}$  be any set of  $N$ -dimensional stochastic vectors. The linear subspace  $\mathcal{CS}(\mathcal{Z}, \Pi, P)$  is defined from (1.4) by replacing the single vector  $\alpha$  by the collection of vectors  $\mathcal{Z}$ . In other words,  $\mathcal{CS}(\mathcal{Z}, \Pi, P)$  is the minimal subspace including  $\mathcal{Z}$  that is invariant under  $P$  and  $\Pi_y, y \in \mathcal{Y}$ . The reader will note that for the notations to be consistent,  $\mathcal{CS}(\alpha, \Pi, P)$  should be interpreted as  $\mathcal{CS}(\{\alpha\}, \Pi, P)$ . We choose to drop the braces in case of a single vector to lighten the notations.

The following result may be proved as Theorem 2.2 (the details are omitted).

**Theorem 2.3** *Let  $\mathcal{Z}$  be a set of  $N$ -dimensional stochastic vectors.  $(\varphi(X_n))$  is a  $k$ th-HMC for every  $\alpha \in \mathcal{Z}$  with a transition matrix that does not depend on  $\alpha$  iff*

$$\forall (y_{k-1}, \dots, y_0) \in \mathcal{Y}^k, \quad P(\text{Ker}(V_\varphi) \cap P^{(y_{k-1} \dots y_0)} \mathcal{CS}(\mathcal{Z}, \Pi, P)) \subset \text{Ker}(V_\varphi). \quad (2.6)$$

Note that the previous result is not valid if the transition matrix of  $(\varphi(X_n))$  is allowed to depend on the probability distribution of  $X_0$  selected in  $\mathcal{Z}$ .

## 2.2 Homogeneous Markov property

This subsection is devoted to the homogeneous (1th-order) Markov property of the lumped process  $(\varphi(X_n))$ .

**Definition 2.2** *If the process  $(\varphi(X_n))$ , is an HMC with transition matrix  $\hat{P}$ , then the Markov chain  $(X_n)$ , or its transition matrix  $P$ , are said to be weakly lumpable with the matrix  $\hat{P}$  (w.r.t. the lumping map  $\varphi$ ).*

It is easily seen that  $(\varphi(X_n))$  may be an HMC with a transition matrix which depend on  $\alpha$ . However, it follows from Corollary 2.2, that this transition matrix only depends on  $P$  and the map  $\varphi$  for a broad class of Markov chains (e.g. see Comment 1,[28]).

### 2.2.1 Local characterization

The cone  $\mathcal{C}_1(\hat{P})$  is from (2.2)

$$\mathcal{C}_1(\hat{P}) = \{\beta \geq \mathbf{0} : \forall y \in \mathcal{Y}, \quad [V_\varphi P - \hat{P} V_\varphi] \Pi_y \beta = 0\}. \quad (2.7)$$

Specializing Theorem 2.1 for  $k = 1$ , we get the following criterion of weak lumpability.

**Corollary 2.3** *The process  $(\varphi(X_n))$  is an HMC with transition matrix  $\hat{P}$  iff*

$$\mathcal{CC}(\alpha, \Pi, P) \subset \mathcal{C}_1(\hat{P}).$$

*If the above inclusion holds, then  $(\varphi(X_n))$  is still an HMC with the transition matrix  $\hat{P}$  for any stochastic vector in  $\mathcal{CC}(\alpha, \Pi, P)$  as initial distribution of  $(X_n)$ .*

*A new sufficient condition for weak lumpability.* Let us check that a sufficient condition for weak lumpability is given by

$$\forall y \in \mathcal{Y}, \quad PU_\alpha = PU_{P_\alpha(y)} \text{ (if well-defined)}. \quad (2.8)$$

This relation is equivalent to

$$\forall y_0, y_1, y_2, \quad \Pi_{y_2} P \Pi_{y_1} P \Pi_{y_0} \alpha \propto \Pi_{y_2} P \Pi_{y_1} \alpha.$$

Therefore, we have

$$\mathcal{CC}(\alpha, \Pi, P) = \text{Cone}(\Pi_{y_0} \alpha, \Pi_{y_2} P \Pi_{y_1} \alpha, y_0, y_1, y_2 \in \mathcal{Y}).$$

As in the “only if” part of the proof of Theorem 2.2, we can define a  $M \times M$  stochastic matrix  $\hat{P}$  such that

$$\forall y \in \mathcal{Y}, \quad \Pi_y \alpha \in \mathcal{C}_1(\hat{P}).$$

Next, multiplying to the left Relation (2.8) by  $V_\varphi$ , we find that

$$\forall y_0, y_1 \in \mathcal{Y}, \quad \Pi_{y_1} P \Pi_{y_0} \alpha \in \mathcal{C}_1(\hat{P}).$$

Then, it follows that  $\mathcal{CC}(\alpha, \Pi, P) \subset \mathcal{C}_1(\hat{P})$  and we deduce from Corollary 2.3 that  $(\varphi_n(X_n))$  is an HMC. Condition (2.8) will be useful in Section 4.

A new criterion for weak lumpability is given by Theorem 2.2 with  $k := 1$ . The main interest in this result is to provide a polynomial algorithm to check the weak lumpability property (see Subsection 2.2.2). Notice, the transition matrix of the lumped process has not to be specified to use the criterion.

**Corollary 2.4** *The process  $(\varphi(X_n))$  is an HMC iff*

$$P(\text{Ker}(V_\varphi) \cap \mathcal{CS}(\alpha, \Pi, P)) \subset \text{Ker}(V_\varphi). \quad (2.9)$$

*If the above inclusion holds, then  $(\varphi(X_n))$  is still an HMC with every stochastic vector in  $\mathcal{CS}(\alpha, \Pi, P)$  as initial distribution of  $(X_n)$ .*

Note that Relation (2.9) is just a reformulation of the Property (2.3) with  $k := 1$

$$\forall y \in \mathcal{Y}, \quad P(\text{Ker}(V_\varphi) \cap \Pi_y \mathcal{CS}(\alpha, \Pi, P)) \subset \text{Ker}(V_\varphi).$$

Since  $\mathcal{CS}(\alpha, \Pi, P)$  is  $P$ -invariant, Relation (2.9) has the equivalent form

$$P(\text{Ker}(V_\varphi) \cap \mathcal{CS}(\alpha, \Pi, P)) \subset \text{Ker}(V_\varphi) \cap \mathcal{CS}(\alpha, \Pi, P).$$

That is, the subspace  $\mathcal{CS}(\alpha, \Pi, P) \cap \text{Ker}(V_\varphi)$  is  $P$ -invariant.

*Rogers-Pitman's condition.* Suppose the stochastic vector  $\alpha$  is such that  $\mathcal{CC}(\alpha, \Pi, P) = \text{Cone}(\Pi_y \alpha, y \in \mathcal{Y})$ , or  $\mathcal{CS}(\alpha, \Pi, P) = \text{Span}(\Pi_y \alpha, y \in \mathcal{Y})$ . Condition (2.9) is trivially satisfied since  $\mathcal{CS}(\alpha, \Pi, P) \cap \text{Ker}(V_\varphi) = \{\mathbf{0}\}$ . Then,  $(\varphi(X_n))$  is an HMC from Corollary 2.4. Note that, for any  $y \in \mathcal{Y}$  such that  $\mathbf{1}^T \Pi_y \alpha \neq 0$ ,  $(\varphi(X_n))$  is still an HMC with  $\alpha^{(y)}$  as probability distribution of  $X_0$ . The present assumption corresponds to a well known sufficient condition for weak lumpability to hold, given by Kemeny and Snell [24, p. 136]:

$$\forall y \in \mathcal{Y}, \quad U_\alpha = U_{P\alpha^{(y)}} \text{ (if well-defined) .} \quad (2.10)$$

This condition is clearly stronger than that defined by Relation (2.8).

The matrix-condition (2.10) has been generalized for the class of HMCs with a general state space by Rogers and Pitman [33]. It is based on the following specific condition satisfied by the transition matrix.

**Definition 2.3** *A stochastic matrix  $P$  is called a R-P matrix if there exist a  $N \times M$  stochastic matrix  $\Lambda$  such that*

$$V_\varphi \Lambda = \hat{I} \quad \text{and} \quad P \Lambda = \Lambda V_\varphi P \Lambda. \quad (2.11)$$

We mention an interesting property of a Markov chain  $(X_n)$  with a R-P transition matrix  $P$ . The probability distributions of random variables  $X_n, n = 1, \dots$  may be computed as follows. We deduce from Relation (2.11) that

$$P^n \Lambda = \Lambda (V_\varphi P \Lambda)^n, \quad \forall n \geq 1.$$

Thus, with any stochastic vector in the cone  $\Lambda \mathbb{R}_+^M$  as probability distribution of  $X_0$ , the one-dimensional distributions of the Markov chain  $(X_n)$  can be computed from the  $M \times M$  matrix  $V_\varphi P \Lambda$ . When  $\Lambda = U_1$ , such a fact is known from [14] and is used in [38, 10]. This was one of the main motivations to deal with R-P matrices for investigating Markov bounds for functions of an HMC in [29].

Specializing Theorem 2.3 for  $k = 1$ , we obtain the following statement.



**Corollary 2.5** *Let  $\mathcal{Z}$  be a subset of probability distributions on  $\mathcal{X}$ .  $(\varphi(X_n))$  is an HMC for every  $\alpha \in \mathcal{Z}$  with a transition matrix that does not depend on  $\alpha$  iff*

$$P(\text{Ker}(V_\varphi) \cap \mathcal{CS}(\mathcal{Z}, \Pi, P)) \subset \text{Ker}(V_\varphi). \quad (2.12)$$

*Strong lumpability.* When  $\mathcal{Z}$  is the set of all the probability distributions over  $\mathcal{X}$  in Corollary 2.5, we obtain the so-called lumpability property or *strong lumpability* property of the transition matrix  $P$ . This property is widely used in stochastic modeling because it can be easily checked on the transition matrix or on the associated graph (e.g. see [38, 19, 9]). Since  $\mathcal{Z}$  is assumed to be the set of  $N$ -dimensional stochastic vectors, we have  $\text{Span}(\mathcal{Z}) = \mathbb{R}^N$ . Since  $\mathcal{Z} \subset \mathcal{CS}(\mathcal{Z}, \Pi, P)$  by definition, it follows that  $\mathcal{CS}(\mathcal{Z}, \Pi, P) = \mathbb{R}^N$ . Then, Relation (2.12) gives the following criterion for strong lumpability:

$$P\text{Ker}(V_\varphi) \subset \text{Ker}(V_\varphi). \quad (2.13)$$

We can take  $\hat{P} := V_\varphi P U_1$  as the transition matrix of  $(\varphi(X_n))$ .

In fact, the following criteria for strong lumpability may be easily derived [24, 6, 12].

**Theorem 2.4** *Let  $P$  be a stochastic matrix and  $\hat{P}$  be the matrix  $V_\varphi P U_1$ . The following statements are equivalent.*

1. *The process  $(\varphi(X_n))$  is an HMC with transition matrix  $\hat{P}$  for every initial distribution of  $(X_n)$ , and  $P$  is said to be strongly lumpable into  $\hat{P}$*
2.  $\forall y_1, y_2 \in \mathcal{Y}, [\sum_{x_2 \in \varphi^{-1}(y_2)} P(x_2, x_1) \text{ does not depend on } x_1 \in \varphi^{-1}(y_1)]$
3.  $V_\varphi P = \hat{P} V_\varphi$
4.  $P\text{Ker}(V_\varphi) \subset \text{Ker}(V_\varphi)$
5.  $\text{Ker}(V_\varphi) \subset \text{Ker}(V_\varphi P)$ .

### 2.2.2 An algorithm for checking the weak lumpability property

Let us outline a finite algorithm to check that  $(\varphi(X_n))$  is an HMC when  $X_0$  has probability distribution  $\alpha$ . Since  $\mathcal{CS}(\alpha, \Pi, P)$  is the minimal subspace that is invariant under  $P$  and  $\Pi_y, y \in \mathcal{Y}$  and that contains  $\text{Span}(\Pi_y \alpha, y \in \mathcal{Y})$ , an algorithm for the evaluation of  $\mathcal{CS}(\alpha, \Pi, P)$  may be designed from [6, p. 209] as follows. Note that the computation of  $\mathcal{CS}(\alpha, \Pi, P)$  allows to check the  $k$ th-order Markov property of  $(\varphi(X_n))$  from Theorem 2.2.

Write the given transition matrix  $P$  in a block form, based on the partition  $\cup_{y \in \mathcal{Y}} \varphi^{-1}(y)$  of  $\mathcal{X}$ , i.e.  $P = \{P_{y_1, y_2} : y_1, y_2 \in \mathcal{Y}\}$ . The same block form is used for the stochastic vector  $\alpha = \{\alpha_y, y \in \mathcal{Y}\}$ . Define the following iterative process: for every  $y \in \mathcal{Y}$

$$\begin{aligned}\mathcal{L}_y^{[0]} &:= \text{Span}(\alpha_y) \\ \mathcal{L}_y^{[n+1]} &:= \sum_{y_1 \in \mathcal{Y}} P_{y, y_1} \mathcal{L}_{y_1}^{[n]} + \mathcal{L}_y^{[n]} \quad n \geq 0.\end{aligned}\tag{2.14}$$

Set  $d := \max_{y \in \mathcal{Y}} (\text{card}(\varphi^{-1}(y)))$ . Since  $\mathcal{L}_y^{[n]} \subset \mathcal{L}_y^{[n+1]}$  and  $\dim(\mathcal{L}_y^{[n]}) \leq \text{card}(\varphi^{-1}(y)) \leq d$ , we easily find that: for all  $y \in \mathcal{Y}$

$$\mathcal{L}_y^{[n]} = \mathcal{L}_y^{[d]} \quad \text{for } n \geq d - 1.$$

Any linear subspace  $\mathcal{L}$  of  $\mathbb{R}^{\text{card}(\varphi^{-1}(y))}$  may be identified with the subspace  $\text{Im}(\mathcal{L}) := \{\mathbf{0}\} \times \mathcal{L} \times \{\mathbf{0}\}$  of  $\mathbb{R}^N$ . That is, any  $\text{card}(\varphi^{-1}(y))$ -dimensional vector  $u \in \mathcal{L}$  is identified with the  $N$ -dimensional vector  $\tilde{u}$  defined by:  $\tilde{u}(x) := 0$  if  $x \notin \varphi^{-1}(y)$  and  $\tilde{u}(x) := u(i)$  if  $x \in \varphi^{-1}(y) = \{x_1, \dots, x_{\text{card}(\varphi^{-1}(y))}\}$ . Then, for every  $y \in \mathcal{Y}$ , we have  $\text{Im}(\mathcal{L}_y^{(n)}) = \Pi_y \mathcal{CS}(\alpha, \Pi, P)$  for all  $n \geq d - 1$ , or

$$\oplus_{y \in \mathcal{Y}} \text{Im}(\mathcal{L}_y^{(n)}) = \mathcal{CS}(\alpha, \Pi, P), \quad n \geq d - 1.$$

Thus, the algorithm (2.14) is finite. More important, the algorithm is polynomial in the number  $N$  of states since it only involves computation of the sum and range of linear subspaces (with “small” dimension). Indeed, at step  $n + 1$ , it is clear from (2.14) that we have to compute the sum of the two linear subspaces

$$\mathcal{L}_y^{[n]} \quad \text{and} \quad \sum_{y_1 \in \mathcal{Y}} P_{y, y_1} \mathcal{L}_{y_1}^{[n]}.$$

Suppose that a basis of  $\mathcal{L}_{y_1}^{[n]}$  for  $y_1 \in \mathcal{Y}$  is given. The main computational task is the computation of the  $M < N$  ranges  $P_{y, y_1} \mathcal{L}_{y_1}^{[n]}$ . Such computations may be performed with the Gaussian elimination procedure which is known to be polynomial in the dimensions and bit-sizes involved (e.g. [17, p. 112]).

In the same way, we get a polynomial-time algorithm to construct  $\mathcal{CS}(\mathcal{Z}, \Pi, P)$ , where  $\mathcal{Z}$  is a set of  $N$ -dimensional stochastic vectors, provided that the minimal linear subspace containing  $\mathcal{Z}$  has “effective” representation. Clearly, in this case, we also can check in polynomial-time the weak lumpability property with respect to the set of initial distributions  $\mathcal{Z}$ . We will use this observation in Section 4.

The next result follows from Corollary 2.4.

**Property 1** *The homogeneous Markov property of  $(\varphi(X_n))$  is checked from  $k \leq \max_{y \in \mathcal{Y}} (\text{card}(\varphi^{-1}(y)))$  steps of the algorithm above.*

Rosenblatt's algorithm to check the weak lumpability is based on Corollary 2.3, which is essentially the criterion of weak lumpability given by Kemeny and Snell [24] for an irreducible matrix  $P$  (see also [35, Th 3.4]). Proposition 1 shows that the cone  $\mathcal{CC}(\alpha, \Pi, P)$  may be computed in at most  $\max_{y \in \mathcal{Y}} (\text{card}(\varphi^{-1}(y)))$  steps. Thus, Rosenblatt's algorithm is finite. However, the extremal rays of cones are needed for these algorithms and their computation is exponential in the dimensions involved. Rosenblatt's algorithm is essentially a Bayesian computation and is "mildly" non-linear. The main contribution of the algorithm associated with (2.14) is to show that it is possible to avoid the "Bayesian" nonlinearity.

**Comment 2** In this paper, we only deal with discrete-time Markov chains. The main criteria for a function of a discrete-time Markov chain to be an HMC carry over in the continuous-time context. Indeed, the "uniformization procedure", which is a basic tool for the numerical analysis of continuous time Markov models [41], may be used to derive from a continuous-time Markov chain  $(X_t)_{t \in \mathbb{R}_+}$ , a discrete-time Markov chain to which our results apply. This discrete-time Markov chain is called the "uniformized" chain associated with  $(X_t)_{t \in \mathbb{R}_+}$ . The following result may be proved (see [22] for further details). For a Markov process  $(X_t)_{t \in \mathbb{R}_+}$  and a lumping map  $\varphi$ , checking that the process  $(\varphi(X_t))_{t \in \mathbb{R}_+}$  is an HMC with a fixed probability distribution of  $X_0$ , consists in applying the algorithm of Subsection 2.2.2 to the uniformized chain associated with  $(X_t)_{t \in \mathbb{R}_+}$ . Under specific assumptions on the continuous time Markov chain  $(X_t)_{t \in \mathbb{R}_+}$ , this property is known from [36, 26].  $\triangle$

### 2.2.3 Global characterization

Many papers are concerned with the derivation of the set  $\mathcal{D}_M(\hat{P})$  of all probability distributions of  $X_0$  such that  $(\varphi(X_n))$  is an HMC with the transition matrix  $\hat{P}$  (e.g. see [35, 27] and the references therein). From Corollary 2.3, this set is the collection of all stochastic vectors in

$$\mathcal{C}_M(\hat{P}) := \{\alpha \in \mathbb{R}_+^N : \mathcal{CC}(\alpha, \Pi, P) \subset \mathcal{C}_1(\hat{P})\}. \quad (2.15)$$

The following properties of the set  $\mathcal{C}_M(\hat{P})$  are easily seen from its definition.

- (P1)  $\mathcal{C}_M(\hat{P})$  is a closed convex cone.
- (P2)  $\mathcal{C}_M(\hat{P})$  is invariant under the matrix  $P$  and the lumping projectors  $\Pi_y, y \in \mathcal{Y}$ .
- (P3)  $\mathcal{C}_M(\hat{P})$  is the maximal sub-cone of  $\mathcal{C}_1(\hat{P})$  that is invariant under all lumping projectors and matrix  $P$ . In other words,  $\mathcal{D}_M(\hat{P})$  is the maximal subset  $\mathcal{Z}$

of the set of all probability distributions over  $\mathcal{X}$ , such that  $P(\mathcal{CS}(\mathcal{Z}, \Pi, P) \cap \text{Ker}(V_\varphi)) \subset \text{Ker}(V_\varphi)$ .

(P4) If  $\Pi_y \mathcal{C}_M(\hat{P}) = \{\mathbf{0}\}$  then, with  $\mathbb{P}$ -probability 1, the state class  $\varphi^{-1}(y)$  can never be accessed by the Markov chain  $(X_n)$  with any initial distribution in  $\mathcal{C}_M(\hat{P})$ .

Property (P4) leads us to define the notion of essentially weakly lumpable matrix.

**Definition 2.4** *The Markov chain  $(X_n)$ , or its transition matrix  $P$ , are said to be essentially weakly lumpable with the matrix  $\hat{P}$  if  $(\forall y \in \mathcal{Y}, \Pi_y \mathcal{C}_M(\hat{P}) \neq \{\mathbf{0}\})$ .*

Note that a weakly lumpable irreducible matrix  $P$  is essentially weakly lumpable from Comment 1.

Any vector  $\alpha$  in  $\mathcal{C}_M(\hat{P})$  is a solution of the linear equations  $\{\mathbf{0} = [V_\varphi P - \hat{P}V_\varphi]P^{(s)}, s \in \mathcal{Y}^*\}$  (see Corollary 2.3). When  $\mathcal{C}_M(\hat{P})$  is non-trivial, it is clear from Proposition 1 that,

$$\mathcal{C}_M(\hat{P}) = \bigcap_{k=0}^{\max_{y \in \mathcal{Y}} (\text{card}(\varphi^{-1}(y)))-1} \{v \in \mathbb{R}_+^N : \forall s \in \mathcal{Y}^*, \text{lg}(s) \leq k, [V_\varphi P - \hat{P}V_\varphi]P^{(s)}v = \mathbf{0}\}.$$

Therefore, the cone  $\mathcal{C}_M(\hat{P})$  has the central property to be polyhedral. Next, the following criterion of weak lumpability is easily derived from [27, Th 3.4].

**Theorem 2.5**  $\mathcal{C}_M(\hat{P}) \neq \{\mathbf{0}\}$  iff there exist a non-negative  $K \times K$  matrix  $Q$  and a non-negative  $N \times K$  matrix  $U$  ( $1 \leq K \leq N$ ) such that

$$\text{Cone}(U) \subset \mathcal{C}_1(\hat{P}), \quad \text{Cone}(U) = \oplus_{y \in \mathcal{Y}} \Pi_y \text{Cone}(U), \quad PU = UQ \quad (2.16)$$

with  $\text{Cone}(U) := U\mathbb{R}_+^M$  and  $\dim \text{Cone}(U) = K$ . In such a case,  $\text{Cone}(U) \subset \mathcal{C}_M(\hat{P})$ .

For a R-P matrix  $P$ , Conditions (2.11) have to be compared to Conditions (2.16). Indeed, the matrices  $U, \hat{P}$  and  $Q$  in (2.16) may be identified with  $\Lambda, Q = \hat{P} = V_\varphi P \Lambda$  in (2.11) respectively. We see from (2.16) that  $\text{Cone}(\Lambda) := \Lambda\mathbb{R}_+^M$  is invariant under  $P$  and the lumping projectors. We have that  $\text{Cone}(\Lambda) \subset \mathcal{C}_M(\hat{P})$  and  $P$  is essentially weakly lumpable. But the set  $\mathcal{C}_M(\hat{P})$  may be larger than  $\text{Cone}(\Lambda)$ .

Let us denote the spectrum of a matrix  $A$  by  $\sigma(A)$ . The following spectral properties arise from the weak lumpability property and can be proved from [5, 27].

**Corollary 2.6** *If  $P$  is essentially weakly lumpable with matrix  $\hat{P}$ , then we have*

$$\forall y \in \mathcal{Y}, \quad \hat{P}(y, y) \in \sigma(\Pi_y P \Pi_y) \quad \text{and} \quad \sigma(\hat{P}) \subset \sigma(P).$$

### 2.2.4 Duality results

In this subsection, the Markov chain  $(X_n)$  and its dual version w.r.t. the scalar product defined in (2.17) are considered. The Markov property of their respective lumped processes is examined. Our results generalize the theorems Th 6-4.5, Th 6-4.8 in [24], where the transition matrix  $P$  is assumed to be primitive (that is irreducible and aperiodic). Note that the method of derivation is new.

Let  $v$  be a positive vector of  $\mathbb{R}^N$ . The  $N \times N$  diagonal matrix with generic diagonal entry  $v(i)$  is denoted by  $\text{diag}(v)$ . We define a scalar product  $\langle \cdot, \cdot \rangle_v$  on  $\mathbb{R}^N$  by

$$\forall x, y \in \mathbb{R}^N, \quad \langle x, y \rangle_v := x^T \text{diag}(v)^{-1} y. \quad (2.17)$$

**Definition 2.5** *Let  $P$  be a  $N \times N$  stochastic matrix. The adjoint matrix  $P^*$  of  $P$  w.r.t. the scalar product  $\langle \cdot, \cdot \rangle_v$  is defined by*

$$P^* := \text{diag}(v) P^T \text{diag}(v)^{-1}.$$

*We have  $(P^*)^* = P$ .*

The matrix  $P^*$  is stochastic if and only if the vector  $v$  is an eigenvector corresponding to the eigenvalue 1 of  $P$ . If  $P^*$  is stochastic, then  $P^*$  is the well known dual or time-reversed matrix of  $P$  [24, Def 3-5.1]. When  $P^* = P$ , the matrix  $P$  is said to be self-adjoint, or reversible in the Markov chain framework. A self adjoint matrix  $P$  is an instance of a normal matrix, that is,  $P$  satisfies  $PP^* = P^*P$ .

Let  $\mathcal{H}$  be a linear subspace of  $\mathbb{R}^N$ . The space  $\mathbb{R}^N$  is the direct sum

$$\mathbb{R}^N = \mathcal{H} \oplus \mathcal{H}^*. \quad (2.18)$$

where  $\mathcal{H}^* := \{w \in \mathbb{R}^N : \langle w, h \rangle_v = 0, h \in \mathcal{H}\}$  is the adjoint subspace of  $\mathcal{H}$ . Note that  $(\mathcal{H}^*)^* = \mathcal{H}$  and we know that

$$P\mathcal{H} \subset \mathcal{H} \iff P^*\mathcal{H}^* \subset \mathcal{H}^*. \quad (2.19)$$

If  $P$  is a normal matrix, then we have [16, p. 275]

$$P\mathcal{H} \subset \mathcal{H} \iff P\mathcal{H}^* \subset \mathcal{H}^*. \quad (2.20)$$

**Theorem 2.6** *Let  $v$  be a positive stochastic vector. We define the linear subspace  $\mathcal{V} := \text{Span}(v^{(y)}, y \in \mathcal{Y})$ . Then,*

1.  $\mathcal{V}$  is  $P$ -invariant if and only if  $\text{Ker}(V_\varphi)$  is  $P^*$ -invariant.

2. When  $P$  is a normal matrix, we have

$$P(\text{Ker}(V_\varphi) \cap \mathcal{CS}(v, \Pi, P)) \subset \text{Ker}(V_\varphi) \iff P\text{Ker}(V_\varphi) \subset \text{Ker}(V_\varphi) \iff P\mathcal{V} \subset \mathcal{V}.$$

*Proof.* A direct computation shows that  $\mathcal{V}^* = \text{Ker}(V_\varphi)$  and hence,  $\mathcal{V} = (\text{Ker}(V_\varphi))^*$ . Applying Relation (2.19) to  $\mathcal{H} := \mathcal{V}$  gives Statement (1). When the matrix  $P$  is normal, the second equivalence in Statement (2) is just Relation (2.20) with  $\mathcal{H} := \mathcal{V}$ . Next, we have  $\mathcal{V} \subset \mathcal{CS}(v, \Pi, P)$  by definition of the last subspace. This inclusion has the equivalent form

$$\mathcal{CS}(v, \Pi, P)^* \subset \mathcal{V}^* = \text{Ker}(V_\varphi).$$

Moreover, the subspace  $\mathcal{CS}(v, \Pi, P)$  is  $P$ -invariant. The matrix  $P$  is assumed to be normal, so that  $\mathcal{CS}(v, \Pi, P)^*$  is  $P$ -invariant from (2.20). Since  $\mathcal{CS}(v, \Pi, P)^* \cap \text{Ker}(V_\varphi) = \mathcal{CS}(v, \Pi, P)^*$ , we find that

$$P(\text{Ker}(V_\varphi) \cap \mathcal{CS}(v, \Pi, P)^*) \subset \text{Ker}(V_\varphi).$$

The first equivalence in Statement (2) is proved as follows. Suppose that  $P(\text{Ker}(V_\varphi) \cap \mathcal{CS}(v, \Pi, P)) \subset \text{Ker}(V_\varphi)$ . Then, we deduce from (2.18) with  $\mathcal{H} := \mathcal{CS}(v, \Pi, P)^*$  and from the inclusion above, that

$$P\text{Ker}(V_\varphi) \subset \text{Ker}(V_\varphi).$$

The converse statement easily follows from the  $P$ -invariance of  $\mathcal{CS}(v, \Pi, P)$ .  $\blacksquare$

When the matrix  $P^*$  is stochastic, the previous theorem reads as follows.

**Corollary 2.7** *Assume that  $P$  has a positive stochastic eigenvector  $v$  associated with its eigenvalue 1. Then*

1.  $P$  is a R-P matrix if and only if  $P^*$  is strongly lumpable.
2. When the matrix  $P$  is irreducible and normal, the three following statements are equivalent:  $P$  is weakly lumpable;  $P$  is strongly lumpable;  $P$  is a R-P matrix.

*Proof.* This is just a reformulation of Theorem 2.6 once the following comments have been mentioned.  $P$  is a R-P matrix with associated matrix  $\Lambda$  iff the cone  $\Lambda\mathbb{R}_+^M$  is  $P$ -invariant, which is easily checked to be equivalent to the subspace  $\Lambda\mathbb{R}^M$  is  $P$ -invariant. Assuming  $P$  irreducible is equivalent to assuming that  $P$  has a unique stochastic eigenvector associated with the eigenvalue 1 [8, p. 27]. We need this assumption only to derive from Statement (1) in Corollary 2.2, that the inclusion  $P(\text{Ker}(V_\varphi) \cap \mathcal{CS}(v, \Pi, P)) \subset \text{Ker}(V_\varphi)$  holds when  $(X_n)$  is weakly lumpable.  $\blacksquare$

### 2.3 Non-homogeneous Markov property

We are interested in a criterion for  $(\varphi(X_n))$  to be a non-homogeneous Markov chain. A related problem was studied by Kelly [23] under the assumption that  $(X_n)$  was a non-homogeneous Markov chain.

**Definition 2.6** *The process  $(\varphi(X_n))$  is a non-homogeneous Markov chain (NHMC) with the transition matrices  $(\hat{P}_n)_{n \geq 0}$  iff for every  $n \geq 0$  and for every  $(y_{n+1}, \dots, y_0) \in \mathcal{Y}^{n+2}$ , we have*

$$\mathbb{P}\{\varphi(X_{n+1}) = y_{n+1} \mid \varphi(X_n) = y_n, \dots, \varphi(X_0) = y_0\} = \hat{P}_n(y_{n+1}, y_n) \\ \text{if } \mathbb{P}\{\varphi(X_n) = y_n, \dots, \varphi(X_0) = y_0\} > 0.$$

Proceeding as in the homogeneous case, we obtain that  $(\varphi(X_n))$  is an NHMC with the transition matrices  $(\hat{P}_n)_{n \geq 0}$  iff for any  $n \geq 0$  and any  $(y_n, \dots, y_0) \in \mathcal{Y}^{n+1}$

$$[V_\varphi P - \hat{P}_n V_\varphi] P^{(y_n \dots y_0)} \alpha. \quad (2.21)$$

Let us define the following sub-cones of  $\mathbb{R}_+^N$ :

$$\mathcal{C}(\hat{P}_n) := \{ \beta \geq 0 : \forall y \in \mathcal{Y}, [V_\varphi P - \hat{P}_n V_\varphi] \beta = 0 \}, \quad n \geq 0. \\ \left. \begin{aligned} \mathcal{CC}_0(\alpha) &:= \bigoplus_{y \in \mathcal{Y}} \text{Cone}(\Pi_y \alpha) \\ \mathcal{CC}_{n+1}(\alpha) &:= \bigoplus_{y \in \mathcal{Y}} \Pi_y P \mathcal{CC}_n, \quad n \geq 0. \end{aligned} \right\} \quad (2.22)$$

Using the cones  $\mathcal{CC}_n(\alpha)$  and Condition (2.21), the following criterion for  $(\varphi(X_n))$  to be an NHMC may be derived as Theorem 2.1. The details are omitted.

**Theorem 2.7**  *$(\varphi(X_n))$  is an NHMC with the transition matrices  $(\hat{P}_n)_{n \geq 0}$  iff*

$$\mathcal{CC}_n(\alpha) \subset \mathcal{C}(\hat{P}_n), \quad \forall n \geq 0.$$

Assume the above inclusion holds and set the initial distribution of  $(X_n)$  to  $P^{(s)} \alpha / \mathbf{1}^\top P^{(s)} \alpha$  for some  $s \in \mathcal{Y}^*$  with  $\text{lg}(s) = k$ . Then,  $(\varphi(X_n))$  is an NHMC with the transition matrices  $(\hat{P}_{n+k})_{n \geq 0}$ .

The last statement in Theorem 2.7 follows from the inclusion  $\mathcal{CC}_n(P^{(s)} \alpha) \subset \mathcal{CC}_{n+k}(\alpha)$  for any  $s \in \mathcal{Y}^*$  such that  $\text{lg}(s) = k$ . Now, we find a criterion in terms of the linear spaces

$$\mathcal{CS}_n(\alpha) := \mathcal{CC}_n(\alpha) - \mathcal{CC}_n(\alpha) \quad n \geq 0,$$

which does not require the knowledge of matrices  $(\hat{P}_n)_{n \geq 0}$ .

**Theorem 2.8** *The process  $(\varphi(X_n))$  is an NHMC iff*

$$\forall n \geq 0, \quad P(\text{Ker}(V_\varphi) \cap \mathcal{CS}_n(\alpha)) \subset \text{Ker}(V_\varphi). \quad (2.23)$$

*Proof* The proof of the “only if” part is as in that of Theorem 2.2.

The converse statement is proved as follows. In a first step, we define the matrices  $(\hat{P}_n)_{n \geq 0}$  as follows. Let  $n \geq 0$  be fixed. For any  $y \in \mathcal{Y}$  such that  $\Pi_y \mathcal{CC}_n(\alpha) \neq \{\mathbf{0}\}$ , select a vector  $\beta$  in this set. We set

$$\hat{P}_n(., y) := V_\varphi P \beta^{(y)}. \quad (2.24)$$

When  $\Pi_y \mathcal{CC}_n(\alpha) = \{\mathbf{0}\}$ ,  $(\varphi(X_n))$  does not visit the state  $y$  at time  $n$  with  $\mathbb{P}$ -probability 1. Then, the stochastic vector  $\hat{P}_n(., y)$  may be arbitrary chosen.

In a second step, we have to prove that  $\Pi_y \mathcal{CC}_n(\alpha) \subset \mathcal{C}(\hat{P}_n)$  for every  $y \in \mathcal{Y}$ . If  $\Pi_y \mathcal{CC}_n(\alpha) = \{\mathbf{0}\}$  then the inclusion is trivial. If not, we just have to justify that

$$\forall \gamma \in \Pi_y \mathcal{CC}_n(\alpha), \quad V_\varphi P \gamma^{(y)} = \hat{P}_n(., y) = V_\varphi P \beta^{(y)} \quad (\text{from (2.24)}).$$

Since  $\gamma^{(y)} - \beta^{(y)} \in \text{Ker}(V_\varphi)$ , we have  $\gamma^{(y)} - \beta^{(y)} \in \mathcal{CS}_n(\alpha) \cap \text{Ker}(V_\varphi)$ . It follows from the inclusion (2.23), that  $\mathbf{0} = V_\varphi P(\gamma^{(y)} - \beta^{(y)})$  and  $V_\varphi P \gamma^{(y)} = V_\varphi P \beta^{(y)}$ .  $\blacksquare$

### Example 2.9

Let us consider an HMC  $(X_n)$  with transition matrix  $P$  and the lumping map  $\varphi$  defined by

$$P = \left( \begin{array}{c|cc} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{array} \right), \quad \varphi(1) = 1, \varphi(2) = \varphi(3) = 2.$$

Take  $e_1 = (1, 0, 0)^T$  as initial distribution. Starting in state 1, the path of process  $(\varphi(X_n))$  is  $1, 2, 2, 1, 2, 2, \dots$  with  $\mathbb{P}$ -probability 1. Therefore  $(\varphi(X_n))$  is a Markov chain. The sequence of transitional matrices may be chosen as

$$\begin{aligned} \hat{P}_n &:= \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \text{ if } n \equiv 0 \pmod{3}; & \hat{P}_n &:= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ if } n \equiv 1 \pmod{3}; \\ \hat{P}_n &:= \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \text{ if } n \equiv 2 \pmod{3}. \end{aligned}$$

However, note that two entries of each matrix of the sequence are arbitrary. Indeed, a simple computation shows that

$$\begin{aligned} \mathcal{CC}_n(e_1) &= \text{Cone}(e_1), & \hat{P}_n(., 1) &= (0, 1)^T, \text{ if } n \equiv 0 \pmod{3}, \\ \mathcal{CC}_n(e_1) &= \text{Cone}(e_2), & \hat{P}_n(., 2) &= (0, 1)^T \text{ if } n \equiv 1 \pmod{3}, \\ \mathcal{CC}_n(e_1) &= \text{Cone}(e_3), & \hat{P}_n(., 1) &= (1, 0)^T \text{ if } n \equiv 2 \pmod{3}. \end{aligned}$$



Since  $\text{Ker}(V_\varphi) = \text{Span}((0, 1, -1))$  and  $\mathcal{CS}_n(e_1) \cap \text{Ker}(V_\varphi) = \{\mathbf{0}\}$ , the criterion in Theorem 2.8 is satisfied.  $\triangle$

A natural question is: can a lumped process  $(\varphi(X_n))$  be an NHMC for every probability distribution of  $X_0$ ? The answer is “negative” if we require that all lumped NHMC share the same sequence of transition matrices  $(\hat{P}_n)_{n \geq 0}$ . In fact, the lumped process will be actually a homogeneous Markov chain. Indeed, if  $(\varphi(X_n))$  is an NHMC with transition matrices  $(\hat{P}_n)_{n \geq 0}$  for every initial distribution, we deduce from Theorem 2.7 that  $\mathcal{C}(\hat{P}_0) = \mathbb{R}_+^N$  and we have  $V_\varphi P - \hat{P}_0 V_\varphi = \mathbf{0}$ . We recognize a condition for  $(X_n)$  to be strongly lumpable (see Theorem 2.4). Let  $e_x$  be the  $x$ th vector of the canonical basis of  $\mathbb{R}^N$ . It can be seen that  $(\varphi(X_n))$  may be an NHMC for any initial distribution  $e_x$  ( $x \in \mathcal{X}$ ) without being an NHMC for every initial distribution [22].

Now, we investigate the conditions under which  $(\varphi(X_n))$  is an NHMC with a periodic sequence of transition matrices  $(\hat{P}_n)_{n \geq 0}$ . We just present a criterion for the 2-periodic case. The general situation is quite similar. It is clear from Theorem 2.7 that  $(\varphi(X_n))$  is an NHMC with a 2-periodic sequence of transition matrices  $(\hat{P}_n)_{n \geq 0}$  iff

$$\begin{aligned} & \forall n \geq 0, \quad \mathcal{CC}_{2n}(\alpha) \subset \mathcal{C}(\hat{P}_0) \text{ and } \mathcal{CC}_{2n+1}(\alpha) \subset \mathcal{C}(\hat{P}_1) \\ \text{or} \quad & \mathcal{CC}_e(\alpha) \subset \mathcal{C}(\hat{P}_0) \text{ and } \mathcal{CC}_o(\alpha) \subset \mathcal{C}(\hat{P}_1) \end{aligned} \quad (2.25)$$

with  $\mathcal{CC}_e(\alpha, \Pi, P) := \sum_{n \geq 0} \mathcal{CC}_{2n}(\alpha)$  and  $\mathcal{CC}_o(\alpha, \Pi, P) := \sum_{n \geq 0} \mathcal{CC}_{2n+1}(\alpha)$ . Note that

$$P \mathcal{CC}_e(\alpha, \Pi, P) \subset \mathcal{CC}_o(\alpha, \Pi, P) \text{ and } P \mathcal{CC}_o(\alpha, \Pi, P) \subset \mathcal{CC}_e(\alpha, \Pi, P). \quad (2.26)$$

In such a context, we obtain the following criterion for weak lumpability. The proof is similar to that of Theorem 2.8 and is omitted.

**Theorem 2.10** *The process  $(\varphi(X_n))$  is a 2-periodic NHMC iff*

$$\begin{aligned} & P(\text{Ker}(V_\varphi) \cap \mathcal{CS}_e(\alpha, \Pi, P)) \subset \text{Ker}(V_\varphi) \cap \mathcal{CS}_o(\alpha, \Pi, P) \\ \text{and } & P(\text{Ker}(V_\varphi) \cap \mathcal{CS}_o(\alpha, \Pi, P)) \subset \text{Ker}(V_\varphi) \cap \mathcal{CS}_e(\alpha, \Pi, P) \end{aligned}$$

where

$$\begin{aligned} & \mathcal{CS}_e(\alpha, \Pi, P) := \mathcal{CC}_e(\alpha, \Pi, P) - \mathcal{CC}_o(\alpha, \Pi, P) \\ \text{and } & \mathcal{CS}_o(\alpha, \Pi, P) := \mathcal{CC}_o(\alpha, \Pi, P) - \mathcal{CC}_e(\alpha, \Pi, P). \end{aligned}$$

### 3 $\varphi$ -Observability

In this section, the matrix  $P$  is the transition matrix of the Markov chain  $(X_n)$ . The following linear system can be considered from matrices  $P$  and  $V_\varphi$

$$x_{n+1} := Px_n \quad \text{and} \quad y_n := V_\varphi x_n \quad n \geq 0,$$

where the “state vector”  $x_n$  corresponds to the probability distribution of the random variable  $X_n$ . The “observed vector”  $y_n$  is the probability distribution of the random variable  $\varphi(X_n)$ .

Let us recall that the pair of matrices  $(P, V_\varphi)$  is said to be *observable* if the so-called *non-observable space* is reduced to  $\{\mathbf{0}\}$ , that is (e.g. see [6])

$$\bigcap_{k \in \mathbb{N}} \text{Ker}(V_\varphi P^k) = \{\mathbf{0}\}.$$

In an intuitive sense, the pair  $(P, V_\varphi)$  is observable if the initial state  $x_0$  (probability distribution of  $X_0$ ) can be computed by suitable processing of the observed vectors (probability distributions of  $\varphi(X_n)$ ). The observability concept above, is appropriate for the study of the one-dimensional distributions of the lumped process  $(\varphi(X_n))$ . It is clear from (2.1) that we need a concept of observability adapted to the study of finite-dimensional distributions (or paths) of the lumped process. Such a concept has been introduced by Gurvits for the investigation of the stability of linear inclusions [21] (see also [20]) and by Amari and its co-authors [3] in the context of identifiability of hidden Markov models [3].

Let us define the *non- $\varphi$ -observable space* by

$$\mathcal{NO}_{P,\varphi} := \{v \in \mathbb{R}^N : \forall s \in \mathcal{Y}^*, \quad V_\varphi P^{(s)}v = 0\}. \quad (3.1)$$

It is easily checked that  $\mathcal{NO}_{P,\varphi}$  is the maximal subspace of  $\text{Ker}(V_\varphi)$  that is invariant under all lumping projectors and the matrix  $P$ . The finite generation of this subspace is similar to that of the non-observable space in linear system theory (e.g. see [6, Section 3]).

**Definition 3.1** *A pair of matrices  $(P, V_\varphi)$  is said to be  $\varphi$ -observable if*

$$\mathcal{NO}_{P,\varphi} = \{\mathbf{0}\}.$$

In an intuitive sense, the pair  $(P, V_\varphi)$  is  $\varphi$ -observable if the probability distribution of  $X_0$  can be computed from the knowledge of the finite-dimensional distributions of  $(\varphi(X_n))$  [22]. That any observable pair  $(P, V_\varphi)$  is  $\varphi$ -observable, is clear from their respective interpretations.

### 3.1 Connection between the sets $\mathcal{C}_M(\hat{P})$ and $\mathcal{NO}_{P,\varphi}$

The decomposable cone  $\mathcal{C}_M(\hat{P})$  defined in (2.15) is connected to the space  $\mathcal{NO}_{P,\varphi}$  as follows. Let  $\alpha_y^{(1)}$  and  $\alpha_y^{(2)}$  be two stochastic vectors in  $\Pi_y \mathcal{C}_M(\hat{P})$ . Consider the two Markov chains with the same transition matrix  $P$  and respective initial distributions  $\alpha_y^{(1)}$  and  $\alpha_y^{(2)}$ . They give rise to the same lumped process, which is an HMC with the transition matrix  $\hat{P}$  and the initial distribution  $V_\varphi \alpha_y^{(1)} = V_\varphi \alpha_y^{(2)} = \hat{e}_y$ , where  $\hat{e}_y$  is the  $y$ th vector of the canonical basis of  $\mathbb{R}^M$ . Then, it follows from (2.1) (1.1) and (3.1) that  $\alpha_y^{(1)} - \alpha_y^{(2)} \in \mathcal{NO}_{P,\varphi}$ . Noting that the set  $\mathcal{NO}_{P,\varphi}$  is the direct sum  $\bigoplus_{y \in \mathcal{Y}} \Pi_y \mathcal{NO}_{P,\varphi}$  and that  $\Pi_y(\mathcal{NO}_{P,\varphi}) \cap \mathbb{R}_+^N = \{\mathbf{0}\}$ , the proof of the next result is easily completed from [28, Th3.2].

**Theorem 3.1** *Suppose the cone  $\Pi_y \mathcal{C}_M(\hat{P})$  in (2.15) contains a stochastic vector  $\alpha_y$ . Then,*

$$\Pi_y \mathcal{C}_M(\hat{P}) = [\text{Span}(\alpha_y) \oplus \Pi_y \mathcal{NO}_{P,\varphi}] \cap \mathbb{R}_+^N \quad \text{and} \quad \dim(\Pi_y \mathcal{C}_M(\hat{P})) = \dim(\Pi_y \mathcal{NO}_{P,\varphi}) + 1$$

where  $\dim(\Pi_y \mathcal{C}_M(\hat{P}))$  is the dimension of the affine hull of the set  $\Pi_y \mathcal{C}_M(\hat{P})$ .

When  $P$  is essentially weakly lumpable, there exists a stochastic vector  $\alpha$  such that

$$V_\varphi \alpha > \mathbf{0} \quad \text{and} \quad \mathcal{C}_M(\hat{P}) = [\mathcal{NO}_{P,\varphi} \oplus \text{Span}(\Pi_y \alpha, y \in \mathcal{Y})] \cap \mathbb{R}_+^N.$$

When the pair  $(P, V_\varphi)$  is  $\varphi$ -observable, we obtain the following criterion for the HMC  $(X_n)$  to be essentially weakly lumpable.

**Corollary 3.1** *Let  $\hat{P}$  be a  $M \times M$  stochastic matrix. If the pair  $(P, V_\varphi)$  is  $\varphi$ -observable (observable) then the following statements are equivalent.*

1.  $P$  is essentially weakly lumpable with matrix  $\hat{P}$ .
2.  $\mathcal{C}_M(\hat{P}) = \Lambda \mathbb{R}_+^N$  for some stochastic matrix  $\Lambda$  such that  $V_\varphi \Lambda = \hat{I}$ .
3.  $P$  is a R-P matrix.

*Proof.* Assume that  $\mathcal{NO}_{P,\varphi} = \{\mathbf{0}\}$ . If  $P$  is weakly lumpable with  $\hat{P}$  then it follows from Theorem 3.1 that  $\mathcal{C}_M(\hat{P}) = \Lambda \mathbb{R}_+^M$  with  $\Lambda \hat{e}_y = \alpha^{(y)}$  for any  $y \in \mathcal{Y}$  and  $V_\varphi \Lambda = \hat{I}$ . Property (P3) (p.18) and Theorem 2.5 give that  $P\Lambda = \Lambda \hat{Q}$  for some non-negative matrix  $\hat{Q}$ . In other words,  $P$  is a R-P matrix. The fact that Statement (3) implies Statement (1) is already known (see pages 15,19).  $\blacksquare$

### 3.2 Nowhere density of the set of weakly lumpable matrices

Let  $\mathcal{M}$  be the linear space of real  $N \times N$  matrices and  $\mathcal{S}$  be the set of stochastic matrices. A subset  $\mathcal{H}$  of  $\mathcal{S}$  is said to be *nowhere dense* in  $\mathcal{S}$ , if  $\overline{\mathcal{H}}^\circ = \emptyset$  where  $\overline{\mathcal{H}}$  and  $\mathcal{H}^\circ$  are respectively the closure and the interior of  $\mathcal{H}$  in  $\mathcal{S}$ . Note that a closed subset  $\mathcal{H}$  of  $\mathcal{S}$  is nowhere dense in  $\mathcal{S}$  if its complement set  $\mathcal{H}^c$  is dense in  $\mathcal{S}$ . It is clear that a finite union of nowhere dense closed subsets of  $\mathcal{S}$  is also nowhere dense in  $\mathcal{S}$ .

The two following lemmas are easily proved from [16, p. 62] and [40, Prop 3.3.12] respectively.

**Lemma 3.1** *The set of all positive stochastic matrices is an open subset of  $\mathcal{S}$ , which is dense in  $\mathcal{S}$ .*

**Lemma 3.2** *Let us fix the lumping map  $\varphi$ . The set  $\mathcal{E}_\varphi := \{P \in \mathcal{S} : (V_\varphi, P) \text{ is not observable}\}$  is a closed subset of  $\mathcal{S}$  that is nowhere dense in  $\mathcal{S}$ .*

**Lemma 3.3** *Let us fix the lumping map  $\varphi$ . The set  $\mathcal{RP}_\varphi$  of the R-P matrices is a closed subset of  $\mathcal{S}$  that is nowhere dense in  $\mathcal{S}$ .*

*Proof.* Let  $(P_n)$  be a sequence of elements in  $\mathcal{RP}_\varphi$  converging to  $P$ . Since  $P_n \in \mathcal{RP}_\varphi$ , there exist stochastic matrices  $\Lambda_n$  and  $\hat{P}_n$  with  $P_n \Lambda_n = \Lambda_n \hat{P}_n$ . The sets of stochastic matrices are compact, so we can extract subsequences  $(\Lambda_{n_k})$ ,  $(\hat{P}_{n_k})$  from  $(\Lambda_n)$  and  $(\hat{P}_n)$  that converge to some stochastic matrices  $\Lambda$  and  $\hat{P}$ , respectively. We have

$$P\Lambda \xleftarrow[k \rightarrow +\infty]{} P_{n_k}\Lambda_{n_k} = \Lambda_{n_k}\hat{P}_{n_k} \xrightarrow[k \rightarrow +\infty]{} \Lambda\hat{P}$$

and  $P \in \mathcal{RP}_\varphi$ .

Let  $\hat{e}_y$  be the  $y$ th vector of the canonical basis of  $\mathbb{R}^M$ . Condition (2.10) for  $P$  to be a R-P matrix may be reformulated as

$$\forall y_1, y_0 \in \mathcal{Y}, \quad \Pi_{y_1} P \Pi_{y_0} \Lambda \hat{e}_{y_0} \propto \Lambda \hat{e}_{y_1}.$$

We have to prove that the interior of  $\mathcal{RP}_\varphi$  w.r.t.  $\mathcal{S}$  is empty. In other words, if  $P \in \mathcal{RP}_\varphi$  then we must find a stochastic matrix  $P_\varepsilon$ , as closed to  $P$  as necessary, such that  $P_\varepsilon \notin \mathcal{RP}_\varphi$ . If  $1 < M < N$  then there exists a class  $\varphi^{-1}(y)$  with  $2 \leq \text{card}(\varphi^{-1}(y)) < N$ . We only perturb the entries of  $P$  corresponding to the transition probabilities from states in  $\varphi^{-1}(y_0)$  to states in  $\varphi^{-1}(y)$  for some  $y_0 \neq y$  to obtain  $P_\varepsilon$ . Since we choose  $I_{y_1} P I_{y_1} = I_{y_1} P_\varepsilon I_{y_1}$  for all  $y_1 \in \mathcal{Y}$ , we just have to assert that the perturbations of entries of  $\Pi_y P \Pi_{y_0}$  may be arbitrary small such that matrix  $P_\varepsilon$

is stochastic and vector  $\Pi_y P \Pi_{y_0} \Lambda \hat{e}_{y_0}$  is not proportional to  $\Lambda \hat{e}_y$  for some  $y_0 \neq y$ . It is clear that this can be done, so that  $P_\varepsilon$  is not a R-P matrix.  $\blacksquare$

Corollary 3.1 is reformulated as follows.

*If  $P$  is essentially weakly lumpable w.r.t. the lumping map  $\varphi$ , then either the pair  $(V_\varphi, P)$  is not observable (not  $\varphi$ -observable) or the matrix  $P$  is a R-P matrix.*

Now, we state the main result of this subsection.

**Theorem 3.2** *The set of all weakly lumpable matrices is nowhere dense in  $\mathcal{S}$ .*

*Proof.* Let us fix a lumping map  $\varphi$  from  $\mathcal{X}$ . The set of all essentially weakly lumpable matrices according to  $\varphi$ , denoted by  $\mathcal{EWL}_\varphi$ , is such that

$$\mathcal{EWL}_\varphi \subset \mathcal{RP}_\varphi \cup \mathcal{E}_\varphi.$$

We deduce from Lemmas 3.2, 3.3 that  $\mathcal{RP}_\varphi \cup \mathcal{E}_\varphi$  is a closed and nowhere dense subset of  $\mathcal{S}$ . Since  $\mathcal{X}$  is assumed to be a finite set, the set of lumping maps from  $\mathcal{X}$  is also a finite set. Hence, the set of essentially weakly lumpable matrices is contained in the finite union  $\cup_\varphi (\mathcal{RP}_\varphi \cap \mathcal{E}_\varphi)$  of closed and nowhere dense subsets. The proof is easily completed if we prove that the set of all non-essentially weakly lumpable matrices is also contained in a nowhere dense closed subset of  $\mathcal{S}$ .

Note that the weak lumpability property for a positive stochastic matrix is equivalent to the essential weak lumpability property. Moreover, we deduce from Lemma 3.1, that the set of all non-positive matrices is closed and nowhere dense in  $\mathcal{S}$ . Consequently, the set of all non-essentially weakly lumpable matrices is included in a nowhere dense closed subset of  $\mathcal{S}$ .  $\blacksquare$

## 4 Markov property for a probabilistic function of a Markov chain

### 4.1 Hidden Markov Chains

Let us give an intuitive description of a hidden Markov chain. The mechanism of such a model is as follows (e.g see [34] for further reading). We have a finite set of “states”, say  $\mathcal{X}$ . At each clock time  $n$ , a new state is entered based upon a transition probability distribution which depends on the previous state (the Markov property). After each transition is made, an observation output symbol in  $\mathcal{Y}$  is

produced according to a probability distribution which depends on the current state. This probability distribution is held fixed for the state regardless of when and how the state is entered. A formal definition of a hidden Markov chain is as follows.

**Definition 4.1** *A bivariate homogeneous Markov chain  $((Y_n, X_n))$  with the state space  $\mathcal{Y} \times \mathcal{X}$  is said to be a hidden Markov chain, if its transition matrix  $Q$  satisfies*

$$\begin{aligned} \forall x, x_1 \in \mathcal{X}, \forall y, y_1 \in \mathcal{Y}, \forall n \geq 1 \\ Q((y, x), (y_1, x_1)) &= \mathbb{P}\{(Y_n, X_n) = (y, x) \mid (Y_{n-1}, X_{n-1}) = (y_1, x_1)\} \\ &= \mathbb{P}\{(Y_n, X_n) = (y, x) \mid X_{n-1} = x_1\} \end{aligned} \quad (4.1)$$

$$= \mathbb{P}\{Y_n = y \mid X_n = x\} \mathbb{P}\{X_n = x \mid X_{n-1} = x_1\} \quad (4.2)$$

$$= G(y, x)P(x, x_1). \quad (4.3)$$

for some  $M \times N$ - (resp.  $N \times N$ ) stochastic matrix  $G$  (resp.  $P$ ). The probability distribution of  $(Y_0, X_0)$  is

$$\forall (y, x) \in \mathcal{Y} \times \mathcal{X}, \quad \mathbb{P}\{(Y_0, X_0) = (y, x)\} = G(y, x) \mathbb{P}\{X_0 = x\}.$$

We assume (without loss of generality) that none of the rows of matrix  $G$  is zero. The processes  $(X_n)$ ,  $(Y_n)$  are called the state process and the observed process of the hidden Markov chain, respectively.

It is easily checked from (4.1) that  $(X_n)$  is an HMC with transition matrix  $P$ . Property (4.2) is the so-called “factorization hypothesis” [15, p 1524]. The random variable  $Y_n$  may be thought of as a probabilistic function of the random variable  $X_n$ . Indeed, we have

$$Y_n = \varphi_n(X_n)$$

where  $(\varphi_n)$  is an independent and identically distributed sequence of maps from  $\mathcal{X}$  into  $\mathcal{Y}$ ; the probability distribution of  $\varphi_n$  is specified by  $\mathbb{P}\{\varphi_n(x) = y\} = G(y, x)$  and  $(\varphi_n)$  is independent of  $(X_n)$ . This explains why the process  $(Y_n)$  was early referred to as a probabilistic function of the Markov chain  $(X_n)$  [7].

Spreij investigated the conditions under which  $(Y_n)_n$  is an HMC [37]. We restrict ourselves to this basic question. He used Rubino and Sericola’s formulation [35] of Kemeny-Snell’s criterion for an irreducible state process  $(X_n)$  and a filtering point of view. Here, we need no special assumptions on  $(X_n)$ . In the next subsection, we just express the basic results in terms of the standard parameters of a hidden Markov chain. The results in Section 2 could also be used to deal with the lumpability property studied in [42].

## 4.2 Lumpability of a hidden Markov chain

In this subsection, the stochastic vector  $\alpha$  stands for the probability distribution of  $X_0$ . The HMC  $((Y_n, X_n))$  can be thought of as the following one-dimensional HMC  $(Z_n)$  with state space  $\{1, \dots, NM\}$ :

$$x \in \mathcal{X}, \quad y \in \mathcal{Y}, \quad Z_n = (y-1)N + x \iff (Y_n, X_n) = (y, x).$$

The marginal process  $(Y_n)$  is  $(\Phi(Z_n))$  with the lumping map  $\Phi$  defined by

$$y \in \mathcal{Y}, \quad (\Phi((y-1)N + x) = y, \quad x \in \mathcal{X}).$$

The transition matrix  $Q$  of  $(Z_n)$  has the form

$$Q = \Delta(G)P(\mathbf{1}^T \otimes I_N) \quad \text{with } \Delta(G) := \begin{pmatrix} \text{diag}(G(1, \cdot)) \\ \vdots \\ \text{diag}(G(M, \cdot)) \end{pmatrix},$$

$\text{diag}(G(y, \cdot))$  is the diagonal matrix with the  $y$ th row of  $G$  as diagonal entries, and  $\otimes$  is the Kronecker product of matrices. The probability distribution of  $Z_0$  is  $\Delta(G)\alpha$ .

From Definition 4.1,  $(Y_n)$  is an HMC when  $X_0$  has probability distribution  $\alpha$  if and only if the marginal process  $(Y_n)$  of  $((Y_n, X_n))$  is an HMC with  $(G\alpha, \alpha)$  as probability distribution of  $(Y_0, X_0)$ . Thus,  $(Y_n)$  is an HMC if and only if  $(\Phi(Z_n))$  is an HMC when  $Z_0$  has probability distribution  $\Delta(G)\alpha$ .

The following criterion for  $(Y_n)$  to be an HMC is derived from Corollary 2.5.

**Theorem 4.1** *Let  $\mathcal{Z}$  be a subset of probability distributions over  $\mathcal{X}$ .  $(Y_n)$  is an HMC with the same transition matrix for every probability distribution of  $X_0$  in  $\mathcal{Z}$  iff*

$$Q(\text{Ker}(V_\Phi) \cap \mathcal{CS}(\Delta(G)\mathcal{Z}, \Pi, Q)) \subset \text{Ker}(V_\Phi).$$

When the inclusion above is satisfied for  $\mathcal{Z}$  reduced to a singleton, the hidden Markov chain  $((Y_n, X_n))$  is said to be *weakly lumpable*. If the inclusion holds for  $\mathcal{Z}$  be the set of all stochastic vectors, the hidden Markov chain  $((Y_n, X_n))$  is said to be *semi-strongly lumpable*. The counterpart of the standard conditions for HMCs to be weakly lumpable are now briefly discussed.

*Semi-strong lumpability of a hidden Markov chain.* We know from Theorem 2.4 that  $(\Phi(Z_n))$  is an HMC for every probability distribution of  $Z_0$  iff

$$Q \text{Ker}(V_\Phi) \subset \text{Ker}(V_\Phi). \tag{4.4}$$

In this case, it is clear from the theorem above that the hidden Markov chain  $((Y_n, X_n))$  is semi-strongly lumpable. Let  $e_x$  denote the  $x$ th vector of the canonical basis of  $\mathbb{R}^N$ . Condition (4.4) has the following algebraic form (see Theorem 2.4 and (4.3))

$$\forall x_1, x_2 \in \mathcal{X}, \quad GPe_{x_1} = GPe_{x_2}, \quad (4.5)$$

that is, all columns of the matrix  $GP$  are identical. It can be seen from Theorem 4.1 that a criterion for the hidden Markov chain to be semi-strongly lumpable is

$$\forall y \in \mathcal{Y} : \quad GPe_{x_1} = GPe_{x_2} \text{ whenever } x_1, x_2 \in \{x \in \mathcal{X} : G(y, x) \neq 0\}. \quad (4.6)$$

As shown by the example below, Condition (4.5) is only sufficient in general for Condition (4.6) to hold. But, both conditions are equivalent when  $G$  is positive for instance.

Let us consider the matrices

$$P := \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/3 & 1/3 \\ 1/6 & 5/12 & 5/12 \end{pmatrix} \quad G := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

Since

$$GPe_1 = (1/2 \ 1/4 \ 1/4)^T \quad GPe_2 = (1/2 \ 3/4 \ 3/4)^T,$$

the hidden Markov chain  $((Y_n, X_n))$  is semi-strongly lumpable from (4.6), but  $(Z_n)$  is not strongly lumpable w.r.t.  $\Phi$  from (4.5).

*Rogers-Pitman's condition for a hidden Markov chain.* Condition (2.10) reads as

$$\forall y \in \mathcal{Y}, \quad U_{\Delta(g)\alpha} = U_{Q(\Delta(g)\alpha)(y)} \quad (\text{if well-defined}),$$

and it asserts that  $(Y_n)$  is an HMC. In terms of matrices  $G$  and  $P$ , this condition has the form: for every  $y_1, y_2 \in \mathcal{Y}$

$$\frac{\text{diag}(G(y_1, \cdot))P\text{diag}(G(y_2, \cdot))\alpha}{G^T(y_1, \cdot)P\text{diag}(G(y_2, \cdot))\alpha} = \frac{\text{diag}(G(y_1, \cdot))\alpha}{G^T(y_1, \cdot)\alpha} \quad (\text{if well-defined}).$$

*Spreij's condition.* Assume the following condition be satisfied

$$\forall y \in \mathcal{Y}, \quad QU_{\Delta(g)\alpha} = QU_{Q(\Delta(g)\alpha)(y)} \quad (\text{if well-defined}). \quad (4.7)$$

Then, it follows (see the discussion from (2.8))

$$\mathcal{CC}(\Delta(G)\alpha, \Pi, Q) = \text{Cone}(\Pi_{y_1}\Delta(G)\alpha, \Pi_{y_3}Q\Pi_{y_2}\Delta(G)\alpha, \quad y_1, y_2, y_3 \in \mathcal{Y})$$



and  $(\Phi(Z_n))$  is an HMC with  $\Delta(G)\alpha$  as probability distribution of  $Z_0$ . The following condition for the hidden Markov chain  $((Y_n, X_n))$  to be weakly lumpable is given by Spreij

$$\forall y \in \mathcal{Y}, \quad P(\mathbf{1}^T \otimes I_N)U_{\Delta(g)\alpha} = P(\mathbf{1}^T \otimes I_N)U_{Q(\Delta(g)\alpha)(y)} \text{ (if well-defined).}$$

Left multiplying by the matrix  $\Delta(G)$ , we obtain Condition (4.7).

**Example 4.2 ([37])**

Let us consider the hidden Markov chain  $((Y_n, X_n))$  with associated matrices

$$P_1 = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/4 & 1/3 & 5/12 \\ 1/4 & 1/3 & 5/12 \end{pmatrix} \quad G = \begin{pmatrix} 1/2 & 2/3 & 1/3 \\ 1/2 & 1/3 & 2/3 \end{pmatrix}.$$

This hidden Markov chain is semi-strongly lumpable since all columns of  $GP_1$  are identical. However, contrary to what is reported in [37], there exists a deterministic lumping map w.r.t. which  $P_1$  is weakly lumpable. Indeed,  $P_1$  is a R-P matrix for the map  $\varphi(1) = 1, \varphi(2) = \varphi(3) = 2$ :

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix}^T \quad \hat{P}_1 = \begin{pmatrix} 1/2 & 1/4 \\ 1/2 & 3/4 \end{pmatrix}.$$

Since  $G$  is positive, the HMC  $(Z_n)$  is also strongly lumpable w.r.t. the lumping map  $\Phi(1) = \Phi(2) = \Phi(3) = 1, \Phi(4) = \Phi(5) = \Phi(6) = 2$ .

Now, consider the matrix  $P_2 = P_1^T$ . We get

$$GP_2 = \begin{pmatrix} 19/36 & 35/72 & 35/72 \\ 17/36 & 37/72 & 37/72 \end{pmatrix}.$$

Therefore, the hidden Markov chain with associated matrices  $P_2, G$  is not semi-strongly lumpable. The matrix  $Q$  of  $(Z_n)$  is

$$Q = \left( \begin{array}{ccc|ccc} 1/4 & 1/8 & 1/8 & 1/4 & 1/8 & 1/8 \\ 2/9 & 2/9 & 2/9 & 2/9 & 2/9 & 2/9 \\ 1/18 & 5/36 & 5/36 & 1/18 & 5/36 & 5/36 \\ \hline 1/4 & 1/8 & 1/8 & 1/4 & 1/8 & 1/8 \\ 1/9 & 1/9 & 1/9 & 1/9 & 1/9 & 1/9 \\ 1/9 & 5/18 & 5/18 & 1/9 & 5/18 & 5/18 \end{array} \right).$$

The lumped process  $(\Phi(Z_n))$  is shown to be an HMC with transition matrix

$$\hat{Q} := \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

only for an initial distribution of  $Z_0$  selected in

$$\mathcal{C} := \text{Cone}\left( \left(\frac{1}{3}, \frac{2}{3}, 0, 0, 0, 0\right)^T; \left(\frac{1}{3}, 0, \frac{2}{3}, 0, 0, 0\right)^T; \left(0, 0, 0, \frac{1}{3}, \frac{2}{3}, 0\right)^T; \left(0, 0, 0, \frac{1}{3}, 0, \frac{2}{3}\right)^T \right).$$

In fact, the hidden Markov chain  $((Y_n, X_n))$  is weakly lumpable only for  $\pi_X = (1/3, 1/3, 1/3)^T$  as probability distribution of  $X_0$ . Indeed, it is easily checked that  $\pi_X$  is the only stochastic vector such that  $\Delta(G)\pi_X = \pi_Z \in \mathcal{C}$ .  $\triangle$

## 5 Conclusion

In this paper, we solve a probabilistic question in a pure linear algebra framework. The question addressed here is: under which conditions does a function of a finite Markov chain have still a Markov property? The linear subspaces we introduced to answer to this question are reminiscent of the so-called “geometric approach” in linear control theory [6, 43]. The interest in using such an approach is that we obtain a collection of results probably tedious to derive by means of probabilistic methods.

The initial motivation for such a question, is to obtain a model reduction via an aggregation of some states of a Markov model. We emphasize that this model reduction is exact, in the sense that the aggregated or lumped process has a Markov property which can be exploited to do numerical computations through standard numerical methods for Markov chains. A vast literature exists on model reduction procedures for dynamic systems and, in particular, for Markov models (e.g. see [2, 13, 10] and the references therein). We emphasize that these methods are approximate methods and are well-suited only for specific conditions on the initial model. Finally, we do not intend to claim that our results produce a method of model reduction which is effective for any Markov model. This is not the case. However, there exists a sufficient number of applications for which lumpability leads to a valuable reduction of the computational complexity, to consider that our work may be useful in practice. At least, it should give some insight on the rationale underlying some methods of model reduction.

Finally, we are only concerned with finite Markov chains in this paper. We just say a few words on the countable state space case. Relevant references to

this discussion are [11, 23, 12, 20, 4, 26]. Most results presented here still hold with minor modifications in the statements. Indeed, we have to take care of some topological issues. For instance, the subspace  $\mathcal{CS}(\alpha, \Pi, P)$  will be the minimal *closed* subspace including the vector  $\alpha$  and that is invariant under the matrix  $P$  and the lumping projectors  $\Pi_y, y \in \mathcal{Y}$ . Series like  $\sum_{y \in \mathcal{Y}}$  must be understood as  $l^1$ -sums. An algorithm for computing  $\mathcal{CS}(\alpha, \Pi, P)$  will be infinite in general. However, the instance of R-P matrix reported in [26] shows that finite generation may happen. Results needing topological assumptions on the operators  $P$  and  $\hat{P}$  are clearly related to the statements on spectral properties as in Corollaries 2.2, 2.6. In particular, the connection between the equation  $V_\varphi P = \hat{P} V_\varphi$  and the spectrum of  $P$  and  $\hat{P}$  is studied in [20]. We must also take care of topological issues in the continuous-time context. Indeed, it is clear that if the generator  $A$  is strongly bounded then most results still hold because the reduction to the discrete-time case mentioned in Comment 2 still applies (see [26]). But if  $A$  is not bounded, then precautions are needed. In particular, a criterion for the existence of a closed subspace invariant under  $A$  is studied in [20].  $\varphi$ -observability can be used in the countable case. We do not go into further details here.

## References

- [1] Anderson, B.D.O. The realization problem for hidden Markov models. *Math. Control Signals Systems* **12** (1999), 80–120.
- [2] Antoulas, A.C. Sorensen, D.C. and S. Gugercin S. A survey of model reduction methods for large-scale systems. In *Structured Matrices in Operator Theory, Numerical Analysis, Control, Signal and Image Processing*, Contemporary Mathematics, AMS 2001.
- [3] Amari, S.-I. Ito, H. and Kobayashi, K. Identifiability of hidden Markov information sources and their minimum degrees of freedom. *IEEE Trans. Inform. Theory* **31** (1992), 324–333.
- [4] Ball, F. and Yeo, G.F. Lumpability and marginalisability for continuous-time Markov chains. *J. Appl. Probab.* **19** (1993) 518–528.
- [5] Barr, D.R. and Thomas, M.U. An eigenvector condition for Markov chain lumpability. *Oper. Research* **25** (1977), 1029–1031.

- [6] Basile, G. and Marro, G. *Controlled and Conditioned Invariants in Linear System Theory*, Prentice Hall 1992
- [7] Baum, L.E. and Petrie, T. Statistical inference for probabilistic function of finite set Markov chains. *Ann. Math. Statist.* **37** (1966), 1554–1563.
- [8] Berman, A. and Plemmons, R.J. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, 1979.
- [9] Benoit, A. Brenner, L. Fernades, P. and Plateau, B. Aggregation of Stochastic Automata Networks with Replicas. *Linear Algebra Appl.* **386** (2004) 111–136.
- [10] Buchholz, P. Exact and ordinary lumpability in finite Markov chain. *J. Appl. Probab.* **31** (1994) 59–75.
- [11] Burke, C.J. and Rosenblatt, M. A Markov function of a Markov chain. *Ann. Math. Statist.* **29** (1958), 1112–1122.
- [12] Coxson, P.G. Lumpability and observability of linear systems. *J. Math. Anal. Appl.* **99** (1984), 435–446.
- [13] Dayar, T. and Stewart, J.W. Quasi Lumpability, Lower-bounding coupling Matrices, and Nearly Completely Decomposable Markov Chains. *SIAM J. Matrix Anal. Appl.* **19** (1997), 482–498.
- [14] Delebecque, F. Quadrat, J.P. and Kokotovic P.V. A unified view of aggregation and coherency in networks and Markov chains. *Internat. J. Control* **40** (1984), 939–952.
- [15] Ephraim, Y. and Merhav, N. Hidden Markov Processes. *IEEE Trans. Inform. Theory* **48** (2002), 1518–1569.
- [16] Gantmacher, F.R. *Theory of Matrices* vol 1, New York: Chelsea 1959
- [17] Golub, G.H. and Van Loan C.F. *Matrix computations*, 3thEd, The Johns Hopkins University Press, 1996
- [18] Grinfeld, M. and Knight, P.A. Weak Lumpability in the  $k$ -SAT Problem. *Appl. Math. Letters* **13** (2000), 49–53.
- [19] Gilmore, S. Hillston, J. and Ribaud, M. An Efficient Algorithm for Aggregating PEPA Models. *IEEE Trans. Soft. Eng.* **27** (2001), 449–464.

- [20] Gurvits, L.N. Geometric approach to some problems of control, observation and lumping. Ph.D. Thesis (in Russian) , Gorky University, 1985.
- [21] Gurvits, L.N. Stability of discrete linear inclusion. *Linear Algebra Appl.* **231** (1995), 48–85.
- [22] Gurvits, L.N. and Ledoux, J. Functions of a finite Markov chain. Tech. Rep. LA-UR-01-1976, Los Alamos National Laboratory, Dec 2001.
- [23] Kelly, F.P. Markovian functions of a Markov chain. *Sankhya Ser. A* **44** (1982) 372–379.
- [24] Kemeny, J.G. and Snell, J.L. *Finite Markov chains*. Springer-Verlag, 1976.
- [25] Kienker, P. Equivalence of aggregated Markov models of ion-channel gating. *Proc. R. Soc. Lond. B* **236** (1989) 269–309
- [26] Ledoux, J. On weak lumpability of denumerable Markov chains. *Statist. Probab. Letters* **25**, (1995), 329–339.
- [27] Ledoux, J. A geometric invariant in weak lumpability of finite Markov chains. *J. Appl. Probab.* **34** (1997), 847–858.
- [28] Ledoux, J. and Leguesdron, P. Weak lumpability of finite Markov chains and pseudo-stationarity. *Stoch. Models* **16** (2000), 46–67.
- [29] Ledoux, J. and Truffet, L. Markovian bounds on functions of finite Markov chains. *Adv. in Appl. Probab.* **33** (2001), 505–519.
- [30] Ledoux, J. Linear dynamics for the state vector of Markov chain functions. *Adv. in Appl. Probab.* **36** (2004), 1198–1211.
- [31] Lee, C. P.-C. Golub, G.H. and Zenios, S.A. A Fast Two-stage Algorithm for Computing Page Rank. Tech. report Stanford University SCCM-03-15 (2003).
- [32] Picci, G. Geometric Methods in Stochastic Realization and System Identification. *CWI Quarterly* **9** (1996) 205–240.
- [33] Rogers, L.C.G. and Pitman, J.W. Markov functions. *Ann. Probab.* **9** (1981) 573–582.
- [34] Rabiner, L.R. and Juang, B.H. An Introduction to Hidden Markov Models. *IEEE ASSP Magazine* (1986), 4–16.

- [35] Rubino, G. and Sericola, B. A finite characterization of weak lumpable Markov processes. Part I: The discrete time case. *Stochastic Process. Appl.* **38** (1991), 195–204.
- [36] Rubino, G. and Sericola, B. A finite characterization of weak lumpable Markov processes. Part II: The continuous time case. *Stochastic Process. Appl.* **45** (1993), 115–125.
- [37] Spreij, P. On the Markov property of a finite hidden Markov chain. *Statist. Probab. Letters* **52** (2001), 279–288.
- [38] Schweitzer, P.J. Aggregation methods for large Markov chains, *Mathematical Computer Performance and Reliability* Eds Iazeolla and others, Elsevier-North Holland, 1984
- [39] Simon, H.A. and Ando, A. Aggregation of variables in dynamic systems. *Econometrica* **29** (1961), 111–138.
- [40] Sontag, E.D. *Mathematical Control Theory*, TAM Vol 6 (2th ed) Springer 1998
- [41] Stewart, W.J. *Introduction to the Numerical Solution of Markov chains*, Princeton University, 1994
- [42] White, L.B. Mahony, R. and Brushe, G.D. Lumpable Hidden Markov Models—Model Reduction and Reduced Complexity Filtering. *IEEE Trans. Automat. Control* **45** (2000), 2297–2306.
- [43] Wonham, W.M. *Linear Multivariable Control: a Geometric Approach*, Springer-Verlag 1979